Attrition in Models of Intergenerational Links in Health and Economic Status in the PSID

By John Fitzgerald
Bowdoin College
9700 College Station
Brunswick ME 04011

Email: jfitzger@bowdoin.edu
Phone: 207-725-3593

September 12, 2010
[Preliminary: please ask before citing]

Prepared for:
Conference on "SES and Health across Generations and over the Life Course", at the University of Michigan, Ann Arbor, September 23-24, 2010.

Funding from the Small Grants Program for Research Using PSID Data is gratefully acknowledged. It bears no responsibility for errors or conclusions. Thanks to Matthew Delaney and William Jacob for excellent research assistance.

Attrition in Models of Intergenerational Links in Health and Economic Status in the PSID

Abstract

Estimation of the intergenerational links in health and economic status requires strong data that permits the following of individuals from childhood to adulthood and permits good controls for family background, neighborhood, and environment. The Panel Study of Income Dynamics is well suited for this purpose owing to its long multiple generation time frame and large number of sibling pairs. Family fixed effects or differences across siblings are used in many studies to help control for unobserved background characteristics. The long time frame required for this type of analysis means that selective attrition potentially biases results as the less healthy drop out. Sibling studies require the joint retention of siblings, and this increases attrition because both must respond.

This paper tests the significance of attrition based on health and related factors in the PSID and investigates the impact of attrition on estimates of intergenerational health, education and earnings correlations, with particular emphasis on sibling models. The paper documents the extent of attrition in the PSID, and compares mean characteristics between the PSID and National Health Inventory Survey in various years. Finally it estimates models that predict attrition over time in the PSID and tests for attrition related changes in coefficients in models of the adult outcomes of PSID children. These models estimate the impact of birth weight and parental background on adult outcomes such as health, earnings, and education for children followed into adulthood, including models with family fixed effects. The sample focuses on PSID children age 0 to 16 in 1968.

The paper finds that although health related attrition occurs and affects unconditional means, the weighted PSID maintains its representativeness along key dimensions including health. Simple sibling correlations in outcomes are slightly higher for those who remain in the panel longer, but the differences are not statistically non-zero. For models of intergenerational links with covariates, subsamples of females show no significant effect of attrition on model coefficients. Evidence for males is more mixed, with results for some outcomes and subsamples suggesting that links are stronger for those who remain in the panel longer. But even for males, the evidence is not very robust and differences due to attrition are frequently are not statistically non-zero. Overall, the paper finds little evidence of attrition bias, but some outcomes for males (education and earnings) could benefit from attention to attrition issues.

Attrition in Models of Intergenerational Links in Health and Economic Status in the PSID

I. Introduction and Aims

The Panel Study of Income Dynamics (PSID) has become a premier data set for investigating intergenerational transmission of health and social and economic status. Since the PSID has followed families and descendents since 1968, it provides a long time frame over which analysts can observe family backgrounds for children and adult outcomes for those children. Links between poor child health and low parental income and later adult income and health outcomes are well established. Understanding the nature of these intergenerational correlations in health and poverty is important in order to formulate appropriate policy responses. See Currie (2009) for a recent review. Estimation of the linkages requires strong data that permits individuals to be followed from childhood to adulthood and permits good controls for family background, neighborhood, and environment.

To help control for unobserved links in background characteristics in the estimation of intergenerational correlations, many studies use family fixed effects, or sibling differences. The PSID has strengths here as well owing to its long multiple generation time frame and large number of sibling pairs (e.g. Smith 2008; Haas 2006). Sibling models offer advantages in controlling for unobserved effects but potential disadvantages for measurement error (Griliches 1979). One aspect of sibling models that has been less thoroughly investigated is the role of attrition, and this is one point of departure for this paper.

Attrition by respondents potentially damages the representativeness of a longitudinal survey over time. As is well understood, bias imparted from selective attrition can affect means and distributions of characteristics in the surviving sample as well as regression coefficients and other parameters. Past work has shown that attrition is higher among minorities, the less educated, and those who move. A key finding relevant to this paper is that attrition is also higher for those in ill health (Reynolds, Frank et al. 2005).

Respondents with failing health may be less likely to choose to respond, may move to an institution and not be followed, or may die. Thus, through time, one might expect that the PSID respondents become healthier than a representative cross-sectional sample. Working against this bias, those with poor health are less mobile and thus may be easier to track (Halliday and Kimmitt 2008). In any case, biases due to selective attrition potentially become worse in intergenerational models requiring sibling pairs because both siblings must survive in the panel to provide adult outcome differences.

The impact of parental background and child health on outcomes of children followed to adulthood in the PSID is an active area of research. This paper tests the degree to which attrition in the PSID might bias estimation of intergenerational health and income linkages, with a particular emphasis on sibling models.

II. Background

Intergenerational Models

Health during childhood is a significant determinant of adult outcomes (Currie 2009; Smith 2008;Johnson and Schoeni 2007). A large literature also reveals that parental SES (Social and Economic Status) is a main determinant of child health, and thus health may be a key link in intergenerational correlation in incomes. Children raised by more affluent parents tend to be healthier. The mechanisms by which this occurs are less well understood. High SES parents may purchase more material health inputs, have better access to health care and information, have more education, perhaps have longer time horizons, and may be healthier themselves. Low SES kids may have lower health at birth from poor fetal environment, genetics, or the interaction between the two. There is ample evidence of a gradient of parental SES and child health (Currie 2009;Currie and Hyson 1999; Case, Lubotsky et al. 2002;Currie and Stabile 2003).

Determining the causal path from parental SES to child outcomes is difficult, particularly given the complexity of family background. Unmeasured family factors may cause a correlation between family income and child outcomes that is not causal. For example, poor maternal health could cause poor child health and low family income, or poor child health could lead to reduced maternal employment and low income (Currie, 2009). Paths matter because in the absence of a causal link policies that solely increase incomes may not improve child health. Untangling the mechanisms requires longitudinal measures family income and SES background together with child health.

Rather than trying to directly measure all relevant family background characteristics, studies often use mother fixed effects or differences across siblings to control for measured and unmeasured fixed family traits (Conley and Bennett 2000; Conley and Bennett 2001; Johnson and Schoeni 2007; Smith 2008). Family or mother fixed effects and models that eliminate family effects by use of differences across siblings will be referred to as sibling difference models, hereafter. Sibling difference models are important because otherwise unobserved family influences can produce spurious correlations in outcomes. Sibling models difference out the confounding common permanent income and parent effects.

Yet several problems complicate identification of impacts in sibling models. First, differencing can exacerbate measurement error in health or other variables (Griliches 1979). However, if measurement error is due to a common mother reporting error, then sibling models can potentially help reveal the true relation (Smith 2007). Second, parents may respond to genetic differences in siblings by, for example, compensating to give more resources to a weaker sibling. This would lead to mitigation of the estimated impact of sibling health differences (Currie, 2009). Third, income differences across sibs during childhood may be due to endogenous factors related to their own or parents health (Smith ,2007; Johnson and Schoeni, 2007). Fourth, sibling differences may be affected by health-based attrition if, for example, weaker siblings are less likely to survive or be retained in a sample.

Attrition in the PSID and Health

Past work on attrition in the PSID has found that although there is substantial attrition that affects the unconditional means of many social and economic variables, the PSID survey weights are remarkably good at preserving sample representativeness

(Becketti, Gould et al. 1988; Fitzgerald, Gottschalk, and Moffitt, 1998a; Lillard and Panis 1998; Ziliak and Kniesner 1998). Furthermore, models that condition on a rich set of demographic covariates tend to show little impact of attrition on parameter estimates. Fitzgerald, Gottschalk, and Moffitt (1998b) investigate the role of attrition in intergenerational correlations in income, education, and welfare, and find that the weighted second generation is generally representative, although intergenerational correlations in some variables like education may be sensitive to attrition. But these studies have not directly investigated health.

Several studies have commented on the relation between attrition and health. Based on the PSID, Halliday and Kimmitt (2008) find that less healthy people are more likely to drop out, but also find that less healthy people are less mobile. Meer, Miller et al. (2003) use the PSID to study the link between wealth and health and conclude that attrition is not driving their results. Studies based on surveys other than the PSID have found significant attrition in longitudinal surveys for those in poor health (Reynolds, Frank et al. 2005; Alderson Haddinott, Kinsey, 2006). These papers have not investigated the way attrition affects estimation of sibling models.

III. Attrition in Intergenerational Models

We are interested in the impact of parental characteristics and child background on adult outcomes. A generic model might relate adult outcomes in year *t* to child background during childhood year *s* as follows:

$$y_{fit} = \beta_0 + \beta_1 xc_{fit} + \beta_2 xc_{fis} + \beta_3 xp_{fis} + \alpha_f + \varepsilon_{fit}$$

where y_{fit} is an outcome for child *i* in family *f* as an adult in year *t*

xc_{fit} stands for characteristics of the adult child in year t

xc_{fis} stands for characteristics of the child in childhood year s

xp_{fis} stand for characteristics of the parents in childhood year s, and

α_f stands for family fixed effects.

A selection framework would stipulate that y_t is observed only if the child remains a respondent in year t. Let A^*_{fit} be a latent indicator of attrition propensity in year t

$A^*_{fit} = Z_{fit}\delta + \eta_f + v_{fit}$ and

A_{fit} =1 for non-response in year t when $A^*_{fit}>0$ and

=0 for observed response when $A^*_{fit} <= 0$.

The vector Z could include xc and xp and a fixed family effect η_f, as well as additional identifying variables not included in the structural model (loosely called instruments). Estimation of a fixed effect model requires adult responses for two or more siblings and thus requires the joint event A_{fit}=1 and A_{fjt}=1 for at least two siblings i and j.

Under certain restrictive assumptions, the selection problem can be avoided. For example, assume that ν_{fi} and ν_{fj} are uncorrelated with η_f and the ε's so that all selection takes place by a common family effect due to the correlation of α and η. This would apply if siblings were always observed or not observed together as part of a family unit. In this case, differencing the structural model across siblings to eliminate the fixed effect α results in a residual Δε that is independent of the selection process. Under the assumption that ν and η are independent of ε, that is, that selection operates only through α, differencing eliminates the selection problem (Verbeek and Nijman, 1992).

While this assumption might be tenable in models where siblings always live together at home, it becomes untenable for adult siblings that live apart. For adult outcomes we typically are interested in siblings that will live apart as adults. In that case it is likely that part of selection is idiosyncratic to the sib and this will result in a selection bias after differencing. This more general case is the one pursued in this paper.

III. Approach of this Study

The general literature on selection has proceeded along two main lines. One approach termed "selection on unobservables" makes distributional assumptions on the distributions of the ε and ν, and estimates a correction as a function of the parameters of the attrition process (Heckman (1979)). Identification requires strong functional form assumptions or the existence of a variable that affects the attrition process but is independent of Δε, often called an instrument. In our context, we are unlikely to find good instruments because variables that affect the probability that siblings remain in the sample will likely affect the outcome differences as well.

The second approach has been termed "selection on observables." This assumes that selection is based on a vector of variables z that affects the attrition probability, and is correlated with the $\Delta\varepsilon$ but is not part of the "structural" model. In general, these zs can be lagged values of the ys (Fitzgerald, Gottschalk, Moffitt 1998a, hereafter "FGM"). Under this approach, to obtain consistent estimates of the structural model, the analyst constructs probabilities of sample retention conditional on z and applies them (inverted) as weights in the structural regression, or more generally as weights in the likelihood function for nonlinear models (Wooldridge 2002 p. 587)[1]. In this paper, our primary concern is testing for bias due to selective attrition in intergenerational models involving health. Construction of weights or other solutions is left for future work.

In what follows, prior to tests for attrition, the paper establishes the extent of attrition and shows that it is selective by looking at unconditional distributions of characteristics. As expected, those who remain in the sample tend to be white and/or more advantaged in income and education. Moreover, those retained tend to have better health. This does not establish that intergenerational correlations are necessarily biased.

As a first test, we investigate selection on both unobservables and observables by comparing the distribution of selected variables from the PSID to those from a nationally representative sample from another survey in the same years. We follow Andreski, McGonagle et al. (2007) who compare the PSID and the National Health Interview Survey (NHIS), a repeated national cross-sectional survey. Although they point to differences between the surveys, they argue that NHIS provides a reasonable sample for comparison.

---

[1] That is, the analyst estimates the probability $P(A_{fi}=0|z_{fi})=p_{fi}$ and weights by $1/p_{fi}$. In a sibling difference model, however, the individual weights may be inadequate because sample selection requires that both sibs survive to the adult outcome period. Thus the proper weight requires that we estimate the joint probability $P(A_{fi}=0, A_{fj}=0|z_{fi}, z_{fj})$. These joint weights will differ from the product of the weights if sibling attrition is correlated as is likely. An interesting point here is that in the sibling differenced equation, the *level* of child health or *levels* of parental health or income could be used as relevant zs. That is, these variables do not belong in the differenced equation but potentially correlate with $\Delta\varepsilon$, the sibling differenced error. These would not be suitable "instruments" in a selection on unobservables framework where the zs cannot correlate with $\Delta\varepsilon$.

We find that the data sets are similar in key ways, but unfortunately we cannot assess the similarity of the impacts of child background on adult outcomes in the NHIS because it does not include both.

Assessing the impact of attrition on conditional models of intergenerational correlations is relatively straightforward in the selection on observables framework. As discussed in FGM, if the attrition process is found to be independent of the lagged outcome variables, then this is evidence that we have no selection on observables (for that outcome model). The logic is that lagged outcome variables are likely related to current residuals from the structural model (ε's). Thus finding that lagged dependent variables do not predict attrition suggests that ε's and Z are independent. To implement the test, we estimate sample retention probits to see whether adult outcomes variables of interest are significant predictors of future retention, conditional on measured background. This tests whether lagged dependent variables (adult outcomes measured early) affect the attrition process (later).

In addition, a complementary test investigates how attrition affects coefficients in particular intergenerational regression models. The analysis tests whether coefficients on parental background and child birth weight in models of adult outcomes differ significantly between full samples of respondents and samples reduced by attrition. This test gives a better read on the magnitude of potential attrition problems, but is essentially just an inversion of the model for the retention probit and thus is closely related (FGM). These methods are developed further after discussing data.

IV. Data Issues

To measure intergenerational links, we need data on adult outcomes of children observed earlier in the panel. The primary sample in this paper is all children aged less than 17 present in the original PSID sample in 1968. This cohort, referred to as the child cohort below, will reach ages 18-34 in 1986, the first year when health measures are available for all family members. The cohort ages to 39-55 by 2007. The PSID has collected data on income, earnings, education, demographic information and many other variables since 1968, so that many aspects of parental background for this cohort of children are directly measured. In addition we observe adult outcomes for the child cohort.

This study will follow the lines of Smith (2008), Haas (2006), Johnson and Schoeni (2007) and use commonly employed adult outcome measures: adult self-assessed general health, earnings, and education. Specifically, in the adult outcome models, we measure health for those who are family heads over the period 1986 to 1991. In addition we look at education outcomes (years of education for those age 24 or more by 1991), and earnings outcomes (average labor income excluding farm and business income over the ages 25 to 34). The averaging over time helps limit measurement error (Solon, 1991). Incomes and labor incomes are deflated to 2001 dollars using the GDP consumption deflator, except when indicated as nominal.

Background characteristics will include child birth weight, parental average income during childhood, parental education, parental health, parental age, sibling ages and birth order effects. A number of studies have focused on the role of birth weight and parental income (Currie, 2009)[2]. Although some of these studies suggest that interactions are important, we will use a specification that includes family income, mother education, and child birth weight without interactions as a more generic intergenerational model.

Health data has been collected at various times in the PSID, with general health assessed for heads and wives since 1984. More detailed health questions were asked in 1986 and since 1999. A general health question was asked in 1986 of all family members and this can be used to measure the child's adult health in that year[3]. In 1999, a retrospective health question about childhood health was asked (general health when respondent was less than 17). Smith (2008) and Haas (2006) both use the 1999

[2] For example, Conley and Bennett (2000) use PSID and find income during pregnancy has little effect on birth weights for singleton births after controlling for mother birth weight or family effects. Conley and Bennett (2001) find that low birth weight of mothers and low income at birth interact to produce low birth weight babies. Johnson and Schoeni (2007) use fixed mother effects and find evidence that income during pregnancy and health insurance coverage during pregnancy have positive impacts on child health and into adulthood. They also find that maternal birth weight and income at birth interact to affect child's birth weight.

[3] The PSID question is "Would you say [your/his/her] health in general is excellent, very good, good, fair, or poor?"

retrospective health survey to measure child health. In what follows, the current study uses birth weight as an indicator of child health. Birth weight information was first collected n 1985. Children with low birth weight (less than 5.5 pounds) can be identified for all children in 1985. For the child sample born after 1985, a continuous measure of birth weight is collected in year following the birth.

A note on the matching of children and parents is in order. This paper links parents and children based on annual relation to head codes. Beginning in 1968, I linked children identified as "child of head" to the family head and to the wife if married. This results in links that include stepchildren. This is acceptable if our primary interest is in links due to the environment of the child (family income, mother education, etc.) and not necessarily in genetic links. In later years of the panel, the PSID undertook efforts to identify biological links (e.g. identifying birth mothers in 1983 and obtaining birth histories in 1985 and to present. See the documentation for the Parent Identification File on the PSID website). My match allows me to form links between parents and children where one or both may have dropped out prior to 1983.

The PSID survey includes two subsamples. The nationally representative Survey Research Center (SRC) sample, and an added oversample from low income neighborhoods from the 1968 Survey of Economic Opportunity (SEO) sample. Weights are provided that allow analysts to combine the subsamples to obtain a weighted representative sample. The SEO subsample has been the subject of some concerns about the clarity of its sample frame (Solon, Corcoran et al. 1987; Brown, 1996). In 1997, the PSID underwent a sample redesign and the SEO sample was cut by about two-thirds[4]. For both of these reasons, results from the SRC subsample are emphasized, although some tables include the combined SEO and SRC samples when indicated. In addition, because the analysis requires both child background and adult outcome measures, the child cohort used here necessarily excludes the Latino sample addition in 1990 to 1995 and the 1997 immigrant sample addition.

[4] The core sample reduction in 1997 retained all of the SRC sample and retained black families from the SEO with probability proportionate their 1968 family weight. Non-black families from the SEO were dropped. (Heeringa and Connor, 1999)

V. Attrition by Child Cohort

As noted earlier the PSID has experienced significant attrition. Table 1 shows sample counts and the proportion of children aged 0 to 16 in 1968 who are in responding families in subsequent years. The top panel shows unweighted counts from the combined SEO and SRC samples. Although there are 8104 children in the 1968 cohort, only 7527 are identified as child of head in 1968. Those identified as child of head can be linked to their parent(s) in 1968 and thus will be the basis for much that follows. About 32 percent of these children are retained in the sample in year 2007, but this unadjusted raw indicator of attrition excludes deliberate sample cuts (more below). This table is simply the proportion responding so those who leave are included in any subsequent years in which they return.

The right hand columns of Table 1 distinguish children in household with no other children age 16 or less ("only child"), and those with sibs age 16 or less present in 1968 ("with sibs"). These designations apply only to 1968—that is, an "only child" in 1968 could have a sibling older than 16 who is not counted in this tabulation. The table shows that children from families with more children are slightly less likely to remain in sample. The last column is of particular interest for sibling models because it shows the response rate of children from this cohort for whom there is at least one other responding sib in that year. That is, the prior column shows response rates for children who initially had sibs in 1968, regardless of these sibs' later status, but the last column requires that that at least two children from the original family unit have remained. Such intact sibling groups are needed for family fixed effect models. Only 27 percent of the original children meet this requirement by 2007 when the cohort is aged 39-55.

This top panel gives a misleading impression of the extent of attrition for two reasons. First, the previously mentioned 1997 sample reduction in the SEO sample was intentional as part of a survey redesign and followed known rules. Second, mortality naturally removes some people from the sample. We observe deaths for those in responding family units, but do not observe it for those who left earlier. A potentially more valid measure of the extent of attrition would estimate the sample reduction due to mortality for those who have left the sample and use this as the relevant base sample count. This correction is not pursued in this paper,[5] but the second panel of Table 1 shows response rates for a sample from which I have excluded those who are known to have died, or who were in families who were later dropped in the SEO sample cut in 1997. Response rates are higher with about 40 percent of children of the head responding.

The SRC sample is of particular interest because it was a randomly drawn (clustered) sample that is nationally representative. The last panel of the table shows the response rates for the SRC sample. Response rates are much higher when the less advantaged SEO sample is excluded, even prior to the 1997 sample cut. About 50 percent of the cohort remains in 2007, and about 44 percent of those initially with siblings have a remaining sibling in 2007.

Figure 1 illustrates response rates for all years. The top figure for the combined SRC and SEO sample shows the initial steep drop in sample response in the first year (about 12 percent) and the slower and fairly steady decline thereafter. In 1993/1994 there was a significant recontact effort that increased response rates in the panel. The SEO sample cut in 1997 produced a significant one-year drop. The middle panel excludes those dropped and who are known to have died and the bottom panel shows the SRC only. The line labeled "with present siblings" shows response rates by year for those who have sibs remaining in the sample. In the top panel children with sibs appear to have higher response rates than those with no sibs, but this "single child" effect is an artifact from the SEO sample having larger family sizes and lower response rates. For the SRC sample, the initial 1968 presence of other siblings does not affect retention rates.

Table 2 begins to address how characteristics of the sample members affect attrition. Black children and male children have lower response rates. The initial age of the children appears to matter little. The extent to which characteristics of the remaining respondents differ from those who attrite is developed further in the next section.

VI. Characteristics by Attrition Status

---

[5] See FMG for an example change when this mortality correction is applied is an earlier PSID sample.

The next tables compare pre-attrition characteristics of those who survive in the panel to those who do not. This gives an indication of the nature of the attrition selection based on observable characteristics. This section asks whether the non-attriting sample is different from those attriting, based on characteristics measured in the base year of 1968.

Table 3 shows characteristics of children in 1968 (aged 0-16) and their mothers and/or fathers in 1968. In these tables, the sample is restricted to children of the family head in 1968. Characteristics are weighted by the 1968 person weight to compensate for the oversampling of low-income families in the SEO subsample. The 1968 weight is used because, in this case, we want to see how characteristics change due to attrition and thus do not want to use later year weights that include an attrition adjustment.

The top panel of the tables indicates whether a child in 1968 remained in the panel in year 2007. Results for response up to year 1986 or 1999 are similar[6]. The first two columns show that children who attrite are more likely to come from minority and lower income backgrounds, a finding consistent with the literature. The mothers and fathers of those child tend to have lower education and earnings, are more likely to be non-white and not married, and families who attrite have lower initial incomes.

Sample exits are further separated into three types: exit by death, exit by the SEO sample cut in 1997, or other exits. The other exits include sample non-response, families lost to follow-up after a move-out, etc. If those who die were to be included in the definition of attrition, then attrition along most dimensions would appear to be more selective (that is, those who die differ more from those continuing than those with other exits). In addition, the 1997 SEO cut systematically reduced the sample of low-income whites. For the panels showing whether a person is in or out in 2007, the last column shows characteristics for those dropped by the SEO cut. Sample weights were recomputed post 1997 to compensate for this cut (Heeringa and Connor, 1999).

Table C4 repeats the analysis using the 1986 characteristics from a base sample of those who responded in that year. Since 1986 is the first year for which we have self-

---

[6] Tables available from author.

reported general health measures for all members of the panel, it is a useful base year from which to judge the impact of health on future attrition. In addition, it is the year in which the youngest of the child cohort enters adulthood (those age 0-16 in 1968 are 18-34 in 1986) and we can begin to observe adult outcomes. By 1986, the panel itself is 18 years old and substantial attrition has already taken place. From the table, we observe that those who remain in sample tend to have higher education levels, are more likely to be married, and have higher incomes, consistent with earlier results that the more advantaged are less likely to attrite.

Members of the child cohort who remain in the sample have somewhat better health. About 26 percent of the cohort who remain in 2007 are in the lower three health categories (good, fair, or poor health) compared to 28.5 percent of those who have left not by death or sample cut, and 46 percent of those who are known to have died. Thus attrition appears to be somewhat selective on health for children as they age.

Mothers and fathers, in the panels below, tend to be less healthy due to their older age. They also show a larger difference between those who are observed to remain in the panel to 2007 compared to those that leave, even excluding those known to die. Thus attrition may be selective on health for both parents. The continuation of parents in the survey is of limited interest if we are only interested in the background characteristics of the child and we have sufficient data on parents from the early years of the panel. But the parent/child pair attrition information is useful if we are interested in characteristics of the parents measured later in a child's life (e.g. parental wealth or elder care needs), or, in the case of health measures, if we are interested in years where we can measure both parent and child health[7].

Attrition by Sibling Pairs.

Sibling pair data can be used to eliminate family fixed effects and to calculate sibling correlations that measure intergenerational correlations. The next set of tables display results for sibling pairs. The sample consists of one observation for each sibling pair, hence

[7] The first general health measures for the parents occur in 1984 when health is asked of heads and wives. In these and similar situations, we want to know whether parent child pairs that survive in the panel are representative. Tables available from the author show attrition rates by parents and children.

three siblings generate three pairs, four siblings generate six pairs, etc. Those with no sibling are excluded. The tables have two parts. The top shows the characteristics of the older sibling of the pair broken out by the attrition status of the *pair* indicating whether both remain in 2007, one or the other has become non-response by reason other than death or sample cut, one or the other has died, or one or the other was trimmed by the SEO sample reduction[8]. The bottom part of the table show correlations between the sibs for three outcome variables: a binary indicator for good health in 1986 (excellent or good health on a 5 point scale), labor income averaged over ages 25-34, and years of education at age 24. These correlations computed for each subsample defined by attrition status.

Table 5A shows the characteristics of siblings including both genders by attrition status. Table 5B and 5C show tabulations for male sibs only and female sibs only, respectively. All three tables show that sibs where both remain in the panel in 2007 have somewhat better health in 1986. They also are more likely to be married, have higher education levels, and higher labor income, consistent with earlier tables[9]. For females, we observe less of a difference in labor income. We conclude that sample retention favors the more advantaged, and healthier, but the differences in health are not large in size.

The bottom panels display sibling correlations based on 1986 data. Sibling correlations in outcomes are one measure of the importance of common sibling background and hence reflect intergenerational links (Solon 1991). If those who remain in the panel are more likely to have higher sibling correlations than those who drop out, then we run the risk of overestimating intergenerational correlations using samples selected by attrition. For brothers, in Table 5B, we obtain a sibling correlation on a binary indicator for good health of .23 for the full sample. For the subsample of intact pairs in 2007, the correlation in 1986 health is somewhat higher at .25. For labor income the correlation of .38 in 1986 rises to .41 for those who remain in 2007, and for education the correlation of .43 for the full sample rises to .48 for the subsample that remain in 2007. Thus the correlations rise

---

[8] The groups were made mutually exclusive by assigning the pair to SEO cut if either were SEO cut, then assigning exit by death if either were know to have died, then assigning remainder to non-response to those with neither SEO cut or death occurring.

[9] Results by individual attrition status rather than attrition of either one of the pair are available from the author. Results are similar, but less pronounced.

slightly in the retained subsample, but the difference between the full sample and the sample with both sibs responding in 2007 are not statistically significant.[10] (P values for that test are shown in the table.)

In the sister sample Table 5C, the results are similar. The sister correlations for health are somewhat higher than the brother correlations, with similar correlations for education and lower correlations for earnings. The 1986 correlations are somewhat higher in the selected subsample that survives to 2007, but the differences are not statistically non-zero at conventional levels[11].

In this section we have established that unconditional means for a variety of outcomes differ somewhat between those who remain in the sample and those who leave. Even though unconditional means and distributions of characteristics may differ for respondents and non-respondents, this does not necessarily indicate attrition bias in conditional models because the bias depends on the model under consideration. In future sections we look at specific intergenerational models. The next section takes a broader view and investigates how well the weighted PSID maintains its representativeness over time.

VII. Comparison of PSID and National Health Interview Survey (NHIS)

Weights constructed based on observable characteristics may be used to correct the unconditional means. A test of the adequacy of the "universal" weights calculated by PSID is whether weighted samples over time continue to mirror nationally representative cross sections. FGM compared PSID to CPS samples and found the weighted PSID was remarkably

[10] As discussed in FGM, the appropriate test is that between the full sample and the selected responding sample, and not a comparison of those dropping out and those in the selected responding sample because, in the presence of attrition, the latter two samples are both potentially biased. The correlation coefficients is estimated as a standardized beta from a regression coefficient of one sib's outcome on the other sib's outcome. The significance test is based on a comparison between the full and responding sample for such a regression using a Hausman-type test that allows for robust standard errors (a test based on suest in Stata which computes a combined robust covariance matrix).

[11] If the male and female samples are pooled, the health correlation is .19 in the 1986 sample and .23 for those surviving to 2007, and the difference is significant at a 5 percent level. The differences for labor income and education between the full and selected 2007 sample are not statistically different in the gender pooled sample.

close. But that study did not consider health measures. This section compares PSID and NHIS health data, and thus extends similar work by Andreski, et al. (2007) .

This section first compares the characteristics of the children age 0 to 16 in 1968 with data from the same age cohort of the NHIS. The initial distributions of characteristics in 1968/69 are shown to be similar. Second, it compares the weighted distributions of characteristics across the surveys for the child cohort as it ages from 0-16 in 1968, to age 18 to 34 in 1986, to age 31 to 47 in 1999, and to age 39-55 in 2007. The repeated cross-sections from the NHIS data are assumed to show the distributions of characteristics for a sample not subject to attrition[12]. These are then compared to the same cohort from the PSID using the "universal" sample weights that have been designed in part to compensate for attrition. This section considers characteristics apart from health and the next section considers health measures.

The NHIS data come from the integrated NHIS series maintained by Minnesota Population Center (Minnesota Population Center, 2010). This series provides information on coding differences over time and in cases of differences it often produces variable constructions that have consistent definitions over time. Two data limitations do not allow us to match the exact years. The PSID starts in 1968 and currently has released data through 2007. The NHIS does not have data for 1968 so data from NHIS 1969 will be used. The last year for which we have NHIS data available is 2006. In both cases, even though the years differ by one, the same *age* cohort is compared in each year. (For ease of exposition, I will refer to the years as 1968 and 2007 with the understanding that the NHIS years differ by one.) Tabulations for both data sets are weighted by person weights, or the person weight of the head of household when applicable. The PSID tabulations include the SRC and SEO subsamples, with some tables splitting out the subsamples.

Table A1 shows the characteristics of the heads of the households in which the children aged 0-16 lived in 1968. The racial composition differs across surveys in part due

[12] This ignores that initial period non-interviews occur, with a frequency that may differ across surveys.

to non-comparable definitions[13]. The PSID shows somewhat lower education levels of the heads, somewhat lower proportion married, and somewhat higher proportion working. Importantly for our purposes, the family size and number of children distributions are quite similar. With the possible exception of education, the samples are roughly similar.

In Table A2, the characteristics of the cohort are shown as it ages into adulthood. Changes in race percentage reflect both variation in definitions over time and changes in sample composition. That is, the cross-sectional cohorts in NHIS allows for immigration, but by construction the cohort in PSID does not[14]. This is notable in the lower proportion Hispanic in the PSID. The PSID continues to have somewhat lower education levels than NHIS and somewhat lower marriage levels, but the differences are fairly stable over time. Employment status diverges with PSID showing a fall over the years not seen in NHIS. The indicator for low nominal income (less than 20000) differs with PSID initially showing substantially higher income, but the difference tends to disappear over time[15]. Lastly, family size is similar in 1986, but over time the NHIS shows larger family sizes than PSID. Again, this likely reflects that immigrant families in the NHIS have larger family sizes.

Overall, it appears that the weighted PSID sample maintains its representativeness over time along several key dimensions. Although coding and question differences between the data sets limit comparability and produce some differences in initial levels, the trends in the data appear roughly similar, apart from employment. Trends in health will be explored more systematically in the next section.

VII. Comparison of Health Measures in PSID and NHIS

Andreski et al. (2007) compare responses on health-related questions for the PSID an NHIS for adults in years 1999, 2001,2003 and 2005. They compare responses by samples of adults aged 18 or over in NHIS to heads and wives in PSID. They note that PSID reports somewhat poorer health, as will be verified below, and report several explorations of the

---

[13] In the PSID Hispanic is considered "other" race in 1968. For NHIS, Hispanic is not a race category but there is not a separate Hispanic indicator in 1969.

[14] That is, the PSID weights restore the 1968 distributions. In 1990-1995 PSID added a sample of Latinos, and in 1997 a sample of immigrants. These added samples are excluded from tabulations in this paper in that we do not observe characteristics during childhood together with adult outcomes for these groups.

[15] Nominal income is compared because the NHIS uses nominal income brackets.

difference. They note that specific general health question asked in each year is very similar. They conclude that observed health differences are not due to demographic differences or due to age differences. They further compare a general health question in the PSID, NHIS, and the Health and Retirement Survey (HRS) for a cohort of adults aged 51-61. They find that the PSID and the HRS track closely, with NHIS "being somewhat of an outlier" showing better health. They conclude that although there are differences, the health-related measures in the PSID and NHIS surveys "align fairly closely." This section makes a somewhat different comparison. It compares health responses for those two surveys for a cohort of children in 1968 as they age over time.

As mentioned previously, the year of the first general health question for all PSID respondents is 1986. Table B1 shows a weighted tabulation of general health for years 1986, 1999, and 2007 for the PSID, and 1986, 1999, and 2006 for NHIS. The table reveals that general health in the PSID is poorer than in the NHIS in each year as previously noted by Andreski et al. (2007). Table B1 also shows that within the PSID the SRC subsample is slightly healthier than the combined SEO and SRC. The tabulations follow the same cohort that is age 0-16 in 1968, 18-34 in 1986, 31-47 in 1999, and 38-55 in 2007 for PSID and 2006 for NHIS. Table B1 shows that health declines as the cohort ages. For this 1968 child cohort, the percentage in good or excellent health falls from 74 percent in 1986 to 62 percent in 2006 for the NHIS, and from 69 in 1986 to 56 in 2007 in PSID (SEO+SRC) as the cohort ages. The key distinction is that both attrition and aging affect the health of the PSID cohort, whereas only aging (and cohort) affect the cross-sectional NHIS. Thus the lower level of initial health in the PSID could reflect attrition prior to 1986, among other things. However, the decline in percentage of those in good or excellent health (the age gradient) is about the same. We next explore factors that might reconcile the surveys.

Table B2 shows general health by demographic group. The demographic composition of the PSID is fixed in 1968 and I have excluded new sample members in these tables. To see if varying the demographic composition of the surveys might explain the lower health in PSID, Table B2 shows the percentage in good or excellent health for six sex-race groups. Even within sex-race groups, the lower health of PSID persists.

To standardize further, we use multivariate analysis with the sex-race indicators and age included with indicators for the survey (PSID) and sample time period. For purposes of this exercise, the year 2006 in NHIS is used to compare to 2007 in PSID, and for simplicity both are labeled as 2007. The sample consists of individuals in the original age cohort (0-16 in 1968) from PSID with health measures in 1986, 1999, and 2007 together with NHIS respondents in those same age cohorts with health measured in 1986,1999,2006. Table B3 displays the results. The first column illustrates that the proportion in good/excellent health is about 10 percent lower in the PSID, with no demographic covariates, a point made previously. Health declines in 1999 and 2007 as the cohorts age. The key interaction of the year with the PSID indictor shows that there is no further difference across surveys in 1999 but somewhat less of a health decline in 2007 in the PSID compared to NHIS.

The results in column 2 are conditioned on age and the demographic variables. When we condition on age, the impact of time (survey year) become positive, perhaps due to improved medical technology or knowledge. Conditioning on demographic variables reduces the PSID indicator by half illustrating that demographic differences explain a significant part of lower PSID health. The third and fourth columns show that when the samples are weighted, the PSID effect is smaller and there is no difference in the trend of health for the PSID and NHIS.

Figure 2 shows health age profiles for PSID and NHIS in various years for the cohort aged 0-16 in 1968. The figure plots a lowess smoothing of health residuals from a regression of health on demographic indicators for race and sex (and interactions) and an intercept shift for PSID. This removes the initial difference in levels of health across the surveys and the mean demographic differences. The method then aligns the residuals by subtracting mean differences at age 19 so that all of the lowess lines go through the same point at age 19. This allows us to focus on the age-health profiles themselves. The figure reveals that the age-health profile is very similar across the surveys.

Overall, the analysis suggests that although PSID may have somewhat lower values for self-reported health than NHIS, the change in health over time for the PSID is similar to that in the NHIS for an aging cohort. There is no clear indication that the PSID respondents are becoming relatively less (or more) healthy than a nationally representative sample of the same cohort, given the initial reported health difference. Of course, we have no way of testing the degree to which the initial difference is related to attrition because we lack health measures at the beginning of the PSID.

VII. Adult Outcome Models

A key difficulty in the intergenerational models of interest in this paper is that we do not observe the baseline health variables for childhood such as birth weight until 1985 and we do not observe adult outcomes until the child cohort has aged to adulthood in the panel. The sample thus has been subjected to significant attrition before we observe the adult outcomes of interest. As discussed in FGM, in order to proceed with tests for attrition with this type of intergenerational data we need a strong assumption. We consider a year r part way through the panel (r<t) and consider that as a baseline year (e.g. 1986). We then test whether coefficients change from that point forward as attrition occurs between years r and t. If the structural coefficients change as we restrict the sample to respondents in the post r period, this is evidence of biasing attrition. The converse does not hold: a finding of no coefficient change does not mean that the relationship is unbiased because the biasing effect could have occurred prior to year r. In essence, we make a monotonicity assumption about the attrition bias: attrition in the post-r period has the same impact (say, sign) as attrition in the pre-r period. Thus observing it in the post-r period tells us that bias likely occurred in both periods.

Thus, the choice of base year r involves a tradeoff. An r early in the panel results in a baseline that has been less subjected to attrition and allows a longer post-r follow-up period to observe the effects of attrition on adult outcomes. But given that our interest is in an initial sample of children, an early r means that we have fewer respondents in the child cohort that have reached adulthood. For our purposes we choose 1986-1991 as the baseline because we then have access to child birth weight as well as parental background for the pre-r period, and the we can observe adult outcome variables including health, earnings, and education when the child cohort is in young adulthood. To get better measures of outcomes, we take advantage of multiple years of outcomes and use the period 1986 to 1991 for health, years of education after age 23, and earnings averaged over the ages 25-34.

To reprise our earlier discussion, we first estimate our structural relation in year r for the sample of respondents in year r where we begin to observe adult outcomes:

Baseline $y_{fir} = \beta_0 + \beta_1 xc_{fir} + \beta_2 xc_{fis} + \beta_3 xp_{fis} + \alpha_f + \varepsilon_{fir} \ \ for\ A_{fir} = 0$, $A_{fjr}=0$.

We then re-estimate the same year r relation for the respondents who have survived to time t > r (with $A_{fjt}=0$ and $A_{fjt}=0$). The test is whether the β coefficients on child and parental background change for the selected sample. For example, if the coefficient on parental incomes becomes a stronger predictor of adult health in the selected sample, then this indicates bias. We will test for selection for retention of individuals as well as for retention of groups with siblings present.

Outcome Regressions

We estimate structural models with three outcome variables: bad health (self reported health is fair or poor on a five point scale), educational attainment (years of education for those age 24 or more), and labor income (average labor income for ages 25-34). The primary background variables of interest are average family income when the child was age 0 to 16, mother's education, and child birth weight. The models also condition on child's race/ethnicity, child's age, mother's age, mother's marital status in 1968, and birth order. Models are estimated with and without mother fixed effects.

We first discuss the health model using the SRC sample restricted to male family heads. The dependent variable is a binary indicator with a one for those with fair or poor general health and a zero for those with excellent, very good, or good health. It is estimated using a linear probability model over the five-year window from 1986 to 1991.[16] In Table D1, the low birth weight indicator predicts a large and significant rise of .025 in the probability of poor health as an adult (the mean of dependent variable fair/poor health is about .06). This finding is consistent with Johnson and Schoeni (2007). The average of family income when the child was age 0 to 16 (in 10000s) exerts an insignificant negative effect, conditional on the other covariates. Higher mother's education reduces the chance of poor health (the omitted mother education group is education less than 12 years). The model establishes that mother's education and child's birth weight have significant impacts on the child's health in early to middle adulthood.

The next column shows the same model estimated on the sample of respondents who remain in the sample in 2007. In the more selected sample, we observe a stronger relation between birth weight and bad health, and a small, but more precisely estimated negative impact of family income. The effects for mother's education are reduced. The test for attrition bias tests whether the coefficients in the "full sample" model of column 1 differ from those of the selected sample.[17] Neither the coefficient for birth weight or for family income is found to be significantly different across the two samples.

As an intermediate step in moving toward the mother fixed effect model, the next two columns of the table restrict the sample in each year to those individuals who have brothers also remaining in the sample. When restricted to this subsample, the coefficient on birth weight doubles. Further, it becomes larger in the selected 2007 sample and significantly different from the 1991 sample. This suggests that attrition may be biasing the birth weight to adult health relation and making it appear stronger, but other specifications discussed below produce more mixed results.

---

[16] Earlier work shows that probits produce the same qualitative results. The model was also estimated using the interval regression technique of Johnson and Schoeni (2007) that recodes the five point health index into a 100 point scale with known cut-points. The results are qualitatively similar.

[17] The difference in coefficients across the two models was tested using a test similar to a Hausman test, adapted to robust standard errors (using suest in Stata). As mentioned previously, a simpler test comparing models for those who remain in and those who fall out is improper since attrition would potentially bias the coefficients in both samples.

The final columns show the model estimated with mother fixed effects. Birth weight and average income during childhood can vary across the sibs. The coefficients on birth weight are the same size as those in the brother's sample without fixed effects but the standard errors rise by half and are no longer statistically different from zero. It appears that a difference between the full and selected sample remains in the fixed effect model.[18]

Table D2 presents coefficients on low birth weight and family income for the male SRC sample for the other outcome variables, education and labor income. In the education regression, low birth weight has unstable effects (sometimes positive, sometimes negative) depending on the subsample. Higher average family income during childhood, however, has stable positive effects on adult educational attainment. In the full sample, an increase of $10000 in family income increases education level by about .2 years. The male sibling sample produces similar results, except that the coefficient on income for the selected 2007 sample becomes larger and significantly different from the 1991 sample. Adding mother fixed effects reduces the size of the coefficients and they lose significance.

Turning to the SRC model for male labor income, we measure earnings as a ten-year average for ages 25-34. Low birth weight has a decided negative effect on adult earnings, and impact that becomes larger (in absolute value) when the selected 2007 sample is used and is statistically significant. When mother fixed effects are added, the coefficient on low birth weight remains large and negative, and also is larger (in absolute value) in the selected 2007 sample.

The appendix tables include results for alternative samples of SRC women, and samples for men and women that include the combined SEO and SRC sub-samples for these same dependent variables. Table D3A shows results for SRC women. Family income coefficients are stable and indicate that higher family income while a child increases education and earnings, but not adult health. The results for birth weight are somewhat unstable and often have counterintuitive signs. The test for attrition bias (whether the 1991 and 2007 samples produce statistically different coefficients) does not reveal significant

---

[18] Testing coefficients across the samples with fixed effects and robust standard errors adds a complication and is left for future work.

differences in coefficients between the samples for birth weight or family income and thus attrition looks less problematic for these intergenerational links for females. Results from the combined SRC and SEO sample are shown in Table D3B and D3C for men and women, respectively, and are qualitatively similar to those in the SRC sample.

Overall, the conclusion from the outcome regressions is a bit mixed. Considering all three dependent variables, it appears that low birth weight and family income affect adult outcomes, but only labor income consistently shows strong effects after one includes mother fixed effects. For men, it appears that coefficients are often larger (in absolute value) in the selected 2007 sample for all the dependent variables, although the differences are not often statistically different from zero. For women, few differences appear. Significant differences would indicate selective attrition that potentially biases the coefficients. To confirm the results we turn to an alternative test: whether the outcome variables are significant predictors of subsequent attrition, conditional on the other covariates.

Retention Probits

A more straightforward test for attrition bias is a test that lagged values of the outcome variables predict future attrition, conditional on the other covariates in the model. To complement the analysis above, the lag period in this case in the base period 1986 to 1991. This section shows probits predicting the probability of responding in 2007 given that the person responded in the 1991 interview. In addition, for sibling models, we need at least two siblings to survive to the outcome year for a mother fixed effect outcome regression like those in the last section. Consequently we are interested in the probability of the event that at least two siblings respond in the later year 2007, and estimate probits for that event from the sample of those with siblings present in 1991. As explained previously, a significant coefficient on a (lagged) outcome variable (health, education, earnings) indicates potential attrition bias.

To simplify the interpretation, the outcome variable for health is collapsed into a single indicator that the person experienced poor/fair health at any time during the 1986 to

1991 period. The first two columns of Table E1 show the probits focusing on health for SRC males. The first column shows the results for individuals and the second column shows results for response by at least two brothers in an original family. Poor or fair health in 1991 is not found to be a significant predictor of retention in 2007 for either individuals or for the brother sample after conditioning on the other covariates. The next two columns reveal that education remains a significant predictor of retention for both individual and brother samples. Labor income predicts an individual's later retention, but not retention in the brother only sample. For females shown in Table E2, none of the lagged outcome variables are significant in predicting later attrition. In either subsample, the pseudo R squared is low and few other covariates are significant predictors of attrition[19].

Overall, the results are again a bit mixed. Although models with education and earnings appear to show some potential bias for males, the health model does not. This weakens the conclusion about bias in health models from the last section, but confirms the analysis for education and earnings. For females, the retention probits confirm that attrition does significantly affect outcome models for females: lagged outcome variables do not predict attrition.

Conclusion

The paper began by establishing that substantial attrition has occurred over the long time frame of the PSID, and that this attrition has been selective. Those with lower income, lower education, and worse health tend to become non-respondents more frequently. Nonetheless, a comparison of respondents from the PSID and NHIS shows that the weighted PSID appears maintain its representativeness along key dimensions, attesting to the value of the PSID supplied weights. As for health, although the PSID consistently has somewhat lower responses (worse health) for general health question compared to the NHIS, the NHIS and the PSID produce similar age-health profiles once one allows for the

[19] Although not relevant for testing attrition bias in the outcome models, it is interesting to note that low birth weight, family income, and mother's education do not predict retention, conditional on the other covariates. The only background variable with a stable effect is birth order for the brother sample, and perhaps mother education in the sister sample.

overall lower level in the PSID. This suggests that attrition is not having a substantial effect on these age-health profiles.

Tests for attrition bias in the estimation of outcome models generally do not find strong evidence of bias. In models predicting adult outcomes (health, education, and earnings), one test is whether coefficients on family background and child birth weight change when the sample is restricted to those who remain respondents in later years. We find a pattern in that these intergenerational coefficients are somewhat larger in absolute value in the selected sample, indicating that samples selected by attrition may be producing stronger intergenerational coefficients. But the coefficients are generally not statistically different and the results are sensitive to stratifying by gender. Subsamples of females show few effects of attrition. Retention probits do not show a significant effect of health on retention rates, although higher education and earnings appear to increase retention for males, after conditioning on other covariates.

Results from sibling models with mother fixed effects show patterns similar to those of individuals. It does not appear likely that the added selection of sibling model will produce sharply higher attrition bias, but tests are at this point incomplete and will be completed in future work. Future work will also focus on more complex models of sibling retention, including consideration of reasons for sample exit, and the impacts of using revised weights based on those models.

The specific nature of the models considered here as well as the age range of the sample limit general conclusions. In a number of models, results are unstable across subsamples and results are often imprecise. Nevertheless, given that, overall the paper finds that intergenerational models of parental background and child birth weight impacts on education and earnings outcomes do not appear to be biased by attrition for female samples, although attention perhaps should be given to attrition for male samples for education and earnings outcomes. Intergenerational analyses that predict adult health outcome variables are not likely significantly biased by attrition for either gender, even in sibling models, for the age ranges considered in this paper.

References

Alderman, H., J. Hoddinott, et al. (2006). "Long term consequences of early childhood malnutrition." Oxford Economic Papers 58(3): 450.

Andreski, P., K. McGonagle, et al. (2007). "An Analysis of the Quality of Health Data in the Panel Study of Income Dynamics." Panel Study of Income Dynamics Technical Paper Series **#07-04**.

Becketti, S., W. Gould, et al. (1988). "The Panel Study of Income Dynamics after Fourteen Years: An Evaluation." Journal of Labor Economics **6**: 472-492.

Brown, C. (1996). Notes on the "SEO" or "Census" Component of the PSID. PSID website.

Case, A., D. Lubotsky, et al. (2002). "Economic Status and Health in Childhood: The Origins of the Gradient." American Economic Review 92(5): 1308-1334.

Conley, D. and N. G. Bennett (2000). "IS BIOLOGY DESTINY? BIRTH WEIGHT AND LIFE CHANCES." American Sociological Review **65**(3): 458-467.

Conley, D. and N. G. Bennett (2001). "Birth Weight and Income: Interactions across Generations." Journal of Health & Social Behavior **42**(4): 450-465.

Currie (2009). "Healthy, Wealthy, and Wise: Socioeconomic Status, Poor Health in Childhood, and Human Capital Development." Journal of Economic Literature 47(1): 87.

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

Currie, J. and M. Stabile (2003). "Socioeconomic Status and Child Health: Why Is the Relationship Stronger for Older Children?" American Economic Review 93(5): 1813-1823.

Fitzgerald, J., P. Gottschalk, et al. (1998a). "An Analysis of Sample Attrition in Panel Data: The Michigan Panel Study of Income Dynamics." The Journal of Human Resources **33(2)**: 251-299.

Fitzgerald, J., P. Gottschalk, et al. (1998b). "The Impact of Attrition on the Second Generation of Respondents in the Panel Study of Income Dynamics." The Journal of Human Resources **33(2)**: 300-344.

Griliches, Z. (1979). "Sibling Models and Data in Economics: Beginnings of a Survey." Journal of Political Economy **87**(5): S37-64.

Haas, S. (2006). "Health selection and the process of social stratification: The effect of childhood health on socioeconomic attainment
" Journal of Health and Social Behavior **47**(4): 339-354.

Halliday, T. J. and M. Kimmitt (2008). Selective Migration and Health. IZA Discussion Paper.

Heckman, J. J. (1979). "Sample Selection Bias as a Specification Error." Econometrica **47**(1): 153-161.

Heeringa, S. and J. Connor (1999). 1997 Panel Study of Income Dynamics
Analysis Weights for Sample Families and Individuals. PSID Technical Paper.

Minnesota Population Center INTEGRATED HEALTH INTERVIEW SERIES. University of Minnesota. http://www.ihis.us/ihis/ (Accessed Sept 12, 2010).

Johnson, R. C. and R. Schoeni (2007). The Influence of Early-Life Events on Human Capital, Health Status, and Labor Market Outcomes Over the Life Course. Population Studies Center Research Report. Ann Arbor, University of Michigan.

Lillard, L. A. and C. W. A. Panis (1998). "Panel Attrition from the Panel Study of Income Dynamics: Household Income, Marital Status, and Mortality." Journal of Human Resources **33**(2): 437-457.

Manski, C. F. (1989). "Anatomy of the Selection Problem." The Journal of Human Resources **24**(3): 343.

Meer, J., D. L. Miller, et al. (2003). "Exploring the Health-Wealth Nexus." Journal of Health Economics **22**(5): 713-730.

Reynolds, J., K. Frank, et al. (2005). The Problem of Attrition in Survey Research on Health: Evidence from Ten Longitudinal Surveys. Presented at ASA meetings in Philadelphia PA., Department of Sociology, Florida State University.

Smith, J. P. (2007). "The Impact of Socioeconomic Status on Health over the Life-Course." Journal of Human Resources 42(4): 739-764.

Smith, J. P. (2008). The Impact of Childhood Health on Adult Labor Market Outcomes. UCD Geary Institute Discussion Paper Series, University College Dublin.

Solon, G., M. Corcoran, et al. (1987). "Sibling and Intergenerational Correlations in Welfare Program Participation." NBER Working Paper **No. 2334**.

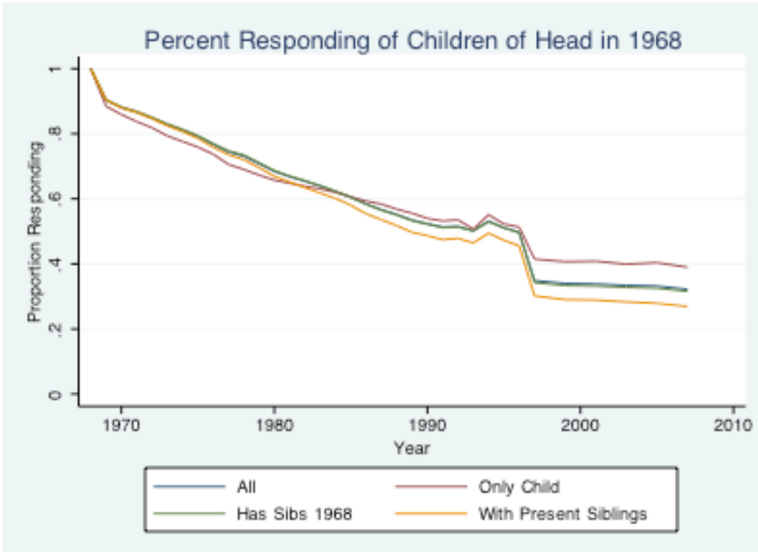
Solon, G. (1991). "A Longitudinal Analysis of Sibling Correlations in Economic Status." Journal of Human Resources 26(3): 509-534.

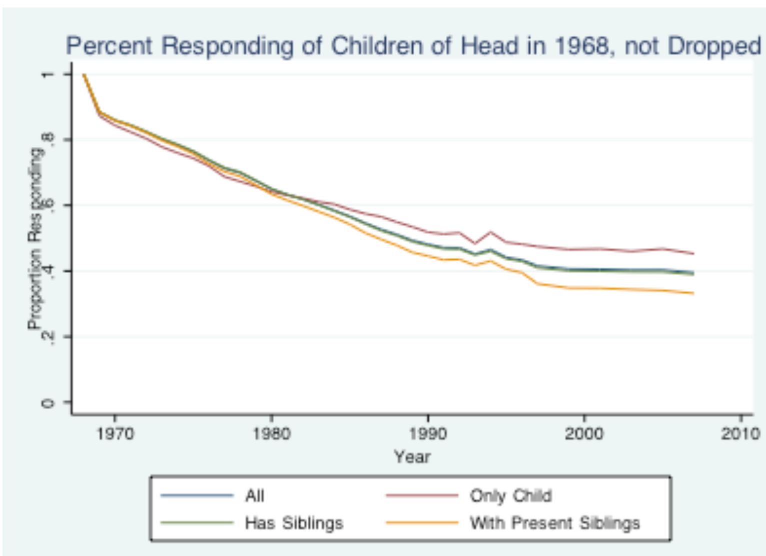
Verbeek, M. and T. Nijman (1992). "Testing for Selectivity Bias in Panel Data Models." International Economic Review 33(3): 681-703.

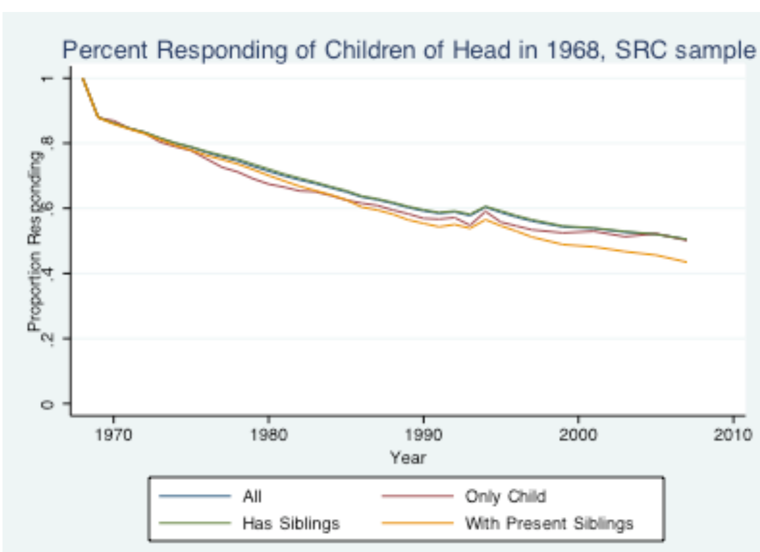
Wooldridge, J. M. (2002). Econometric Analysis of Cross Section and Panel Data. Cambridge MA, MIT Press.

Ziliak, James P., and Thomas J. Kniesner. 1998. "The Importance of Sample Attrition in Life Cycle Labor Supply Estimation." Journal of Human Resources 33, no. 2: 507-530.

Figure 1: Percent Responding in PSID: Children Aged 0 to 16 in 1968

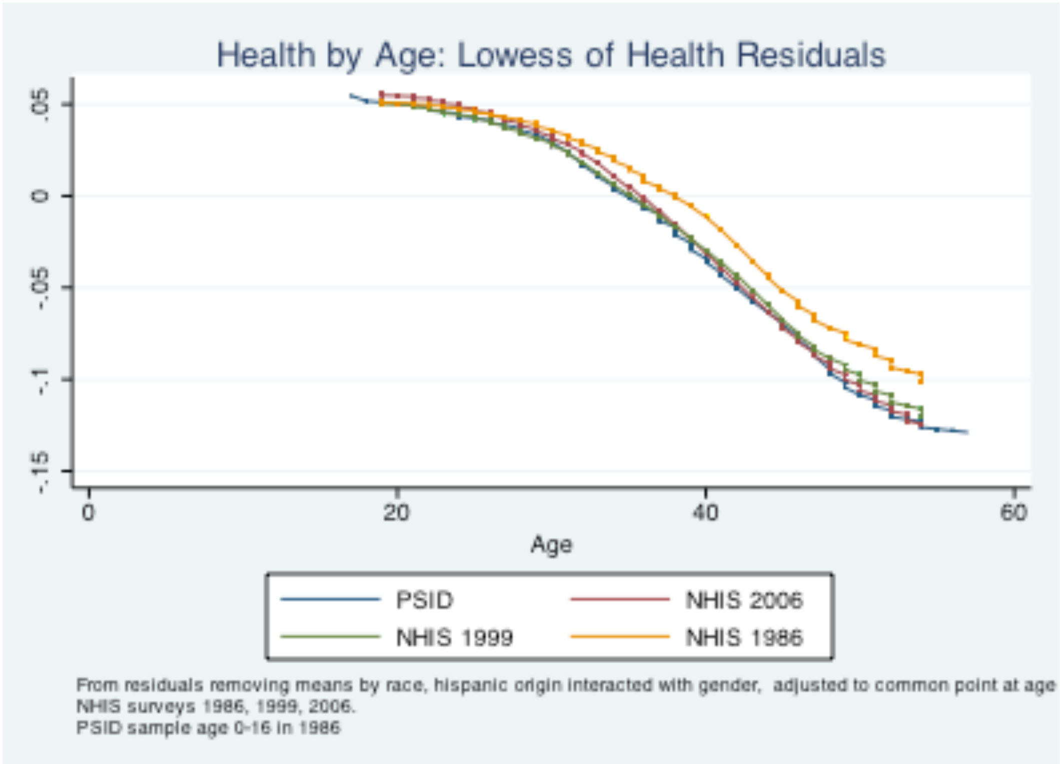






source: author's computation.

Figure 2



Source: author’s computation

Table 1

Proportion of PSID Sample Responding:
Cohort Aged 0-16 in 1968

SRC and SEO Samples

Year	Sample	All	Child of Head	Only child	Has sib in 1968	Presently has Sibs
N		8104	7527	639	6888	6888
1968	8104	1	1	1	1	1
1986	4686	0.578	0.584	0.593	0.583	0.554
1999	2715	0.335	0.340	0.407	0.334	0.29
2007	2576	0.318	0.322	0.390	0.316	0.269

Excluding those cut from SEO Sample or Known Dead

Year	Sample Freq	All %	Child of Head %	Only Child %	Has sib in 1968 %	Presently has Sibs
1968	6623	1	1	1	1	1
1986	3573	0.539	0.546	0.575	0.543	0.492
1999	2644	0.399	0.406	0.465	0.400	0.342
2007	2576	0.389	0.395	0.453	0.389	0.329

SRC ONLY

Year	Sample	All	Child of Head	Only child	Has sib in 1968	Presently has Sibs
N		3279	3108	427	2681	2681
1968	3279	1	1	1	1	1
1986	2050	0.625	0.635	0.616	0.638	0.604
1999	1751	0.534	0.543	0.525	0.546	0.489
2007	1628	0.496	0.504	0.501	0.505	0.435

Notes: Unweighted counts. Responding means in responding family unit or in institution.

Table 2

Proportion Responding by Characteristics of Child
For Children of Head in 1968

A. SRC+SEO Sample

Year	Sample Freq	Child of Head %	White %	Black %	Male %	Female %	Age 0-6 %	Age 7-12 %	Age 13-16 %
1968	7527	1	1	1	1	1	1	1	1
1986	4396	0.584	0.653	0.535	0.556	0.613	0.599	0.585	0.559
1999	2558	0.340	0.420	0.280	0.303	0.378	0.343	0.340	0.335
2007	2424	0.322	0.390	0.274	0.282	0.363	0.322	0.322	0.322

B. SRC + SEO excluding SEO cut and those known dead

Year	Sample Freq	Child of Head %	White %	Black %	Male %	Female %	Age 0-6 %	Age 7-12 %	Age 13-16 %
1968	6136	1	1	1	1	1	1	1	1
1986	3350	0.546	0.617	0.499	0.515	0.577	0.559	0.545	0.525
1999	2493	0.406	0.506	0.328	0.365	0.447	0.400	0.407	0.415
2007	2424	0.395	0.481	0.331	0.353	0.436	0.382	0.397	0.413

C. SRC ONLY

Year	Sample Freq	Child of Head %	White %	Black %	Male %	Female %	Age 0-6 %	Age 7-12 %	Age 13-16 %
1968	3108	1	1	1	1	1	1	1	1
1986	1973	0.635	0.670	0.476	0.616	0.655	0.646	0.625	0.633
1999	1687	0.543	0.583	0.338	0.513	0.575	0.545	0.545	0.537
2007	1567	0.504	0.541	0.319	0.477	0.533	0.501	0.506	0.506

Table 3 (C3)

1968 characteristics by attrition status in 2007

Sample: Children 0-16 in 1968

Variable	All	in 2007	out 2007 (NR, not dead or SEO cut)	out 2007 by death	out 2007 by SEO cut
Charateristics of Child					
Sample	7527	2424	3712	333	1058
Child Race					
White	82.32	88.44	76.04	77.69	86.54
Black	13.86	10.26	18.13	17.95	8.2
Other	3.82	1.3	5.83	4.36	5.26
Mean Family Size	5.74	5.52	5.67	6.42	6.71
Child No. of Sibs					
1 child	12.64	13.65	13.24	12.38	5.6
2 children	23.53	25.92	23.71	16.79	15.08
3 children	23.24	25.82	22.9	16.06	16.56
4 or more	40.59	34.61	40.15	54.77	62.76
Child Gender					
Male	51.14	48.17	52.93	69.62	48.16
Female	48.86	51.83	47.07	30.38	51.84
Family Income					
Mean	41407	46043	39446	37340	31985
Standard Deviation	28373	27433	30991	22107	15643
Median	36580	42992	34325	31415	30124
Characteristics of Mother					
Sample	2995	781	1020	785	409
Marital Status					
Married	86.06	92.59	86.48	78.98	78.43
Widowed	5.44	2.47	3.55	12.15	3.47
Divorced/Separated	7.25	3.83	8.55	7.8	16.24
Never Married	1.25	1.11	1.42	1.07	1.86
Education					
0_11	39.44	24.97	45.66	46.61	54.62
12	44.17	50.75	42.33	40.9	32.45
13_15	9.83	14.9	7.23	6.82	8.52
16+	6.56	9.38	4.79	5.67	4.41
Mean Age	38.07	34.61	35.84	46.14	34.95
Employed	0.4724	0.4854	0.4419	0.4966	0.4592
Labor Income					
Mean	11245	12300	10989	10822	8534
Standard Deviation	9608	10551	8546	9513	8726
Median	9468	10759	10328	8607	6455

Table 3 (C3) (cont.)

Variable	All	in 2007	out 2007 (NR, not dead or SEO cut)	out 2007 by death	out 2007 by SEO cut
Characteristics of Father					
Sample	2283	475	719	877	212
Marital Status					
Married	98.22	99.12	98.64	97.36	97.57
Widowed	0.61	0	0.6	1.03	0.76
Divorced/Separated	1.05	0.46	0.76	1.6	1.64
Never Married	0.12	0.42	0	0.01	0.03
Education					
0_11	40.64	20.41	47.03	49.35	46.57
12	31.11	36.61	28.81	29.53	26.53
13_15	13.55	16.78	13.22	10.71	20.55
16+	14.69	26.2	10.93	10.42	6.35
Mean Age	40.7	35.11	38	47.14	36.81
Employed	0.9679	0.9962	0.976	0.9419	0.9691
Labor Income					
Mean	35373	41413	34087	33266	25082
Standard Deviation	22680	22778	24733	20944	10483
Median	31893	38301	29987	30985	25821

Notes: Sample from PSID. Child of Head, age 0-16 in 1968. Weighted by 1968 person weight. SEO cut is 1997 sample reduction of SEO subsample.

Table 4 (C4)

1986 Characteristics of Children as Adults by Attrition Status in 2007

	All	in 2007	out 2007 (NR, not dead or SEO cut)	out 2007 by death	out 2007 by 97 SEO cut
Charateristics of Adult Child					
Sample	4396	2208	1142	136	910
Health					
Excellent	34.72	36.61	31.1	26.2	34.18
Very Good	37.18	37.78	40.43	27.65	31.24
Good	21.97	20	22.86	27.75	27.97
Fair	5.29	5.08	4.87	13.57	5.19
Poor	0.83	0.54	0.74	4.83	1.43
Marital Status					
Married	68.38	70.51	65.32	55.4	66.66
Widowed	2.74	2.13	2.49	11.81	4.09
Divorced/Separated	11.13	9.78	13.34	10.57	13.48
Never Married	17.75	17.59	18.84	22.22	15.76
Education					
0_11	16.4	13.04	20.29	22.74	23.34
12	40.49	38.69	41.55	52.44	44.12
13_15	24.96	26.85	23.31	19.06	20.68
16+	18.14	21.42	14.85	5.76	11.86
Mean Family Size	3.08	3.05	3.05	2.94	3.28
Mean Age	26.2	26.3	25.78	27.47	26.26
Employed	0.8172	0.8267	0.7993	0.7699	0.8145
Labor Income					
Mean	21876	22772	20405	16915	21211
Standard Deviation	19982	21907	16263	15224	16731
Median	18116	19123	17928	12926	18973
Characteristics of Mother					
Sample	1857	733	247	499	378
Health					
Excellent	14.41	18.76	14.45	6.97	13.55
Very Good	26.11	30.65	26.39	16.52	30.18
Good	34.07	36.86	32.5	31.81	28.37
Fair	17.81	11.39	19.3	27.72	20.1
Poor	7.6	2.34	7.36	16.99	7.8
Marital Status					
Married	72.45	79.84	75.89	60.1	65.29
Widowed	14.59	9.81	6.29	27.05	14.61
Divorced/Separated	12.29	9.86	16.62	12.26	18.99
Never Married	0.67	0.5	1.2	0.59	1.1
Education					
0_11	28.8	18.94	33.13	38.59	45.54
12	45.09	49.1	44.82	40.84	37.04
13_15	14.95	17.87	13.72	12.56	8.46
16+	11.15	14.09	8.33	8	8.97
Mean Age	55.26	52.51	51.75	62.66	52.82
Employed	0.5786	0.6851	0.6218	0.3488	0.6356
Labor Income					
Mean	18202	18905	19164	15574	17163
Standard Deviation	15361	14775	14842	18192	13928

Median	14940	16643	17856	10458	14940

Table 4 (C4) (cont.)

	All	in 2007	out 2007 (not dead)	out 2007 by death	out 2007 by 97 SEO cut
Characteristics of Father					
Sample	1240	436	145	465	194
Health					
Excellent	19.07	28.18	27.55	6.92	14.19
Very Good	27.97	33.93	26.03	22.52	23.96
Good	30.01	27.67	26.65	33.78	29.93
Fair	16.25	9.8	14.77	22.89	22.42
Poor	6.7	0.41	5	13.89	9.49
Marital Status					
Married	91.17	90.92	95.12	92.18	81.26
Widowed	2.43	1.57	0.04	3.56	5.57
Divorced/Separated	6.38	7.51	4.84	4.27	12.89
Never Married	0.02	0	0	0	0.28
Education					
0_11	33.34	22.45	36.11	42.47	45.61
12	30.72	29.13	25.73	34	31.31
13_15	15.03	17.37	17.76	12.19	11.42
16+	20.92	31.06	20.4	11.33	11.67
Mean Age	56.65	52.89	53.26	62.37	55.02
Employed	0.7579	0.895	0.8394	0.576	0.7511
Labor Income					
Mean	46668	57444	44954	31286	34812
Standard Deviation	51030	63475	37649	26178	22364
Median	37960	47807	35108	29773	32867

Notes: Sample from PSID. Child of head, age 0-16 in 1968, age 18-34 in 1986. Weighted by 1968 person weight. SEO cut is 1997 sample reduction of SEO subsample.

Table 5A (C4):
1986 characteristics by attrition status for Sibling Pairs

Variable	All	both in 2007	either out, NR	either out death	either out SEO cut
Charateristics of Child					
Sample	5699	2000	2184	227	1288
Health					
excellent	31.82	33.5	28.27	20.85	35.65
very good	35.92	38.23	39.65	29.27	28.62
good	25.21	22.65	25.26	35.94	27.41
fair	5.88	4.87	5.99	9.1	6.95
poor	1.16	0.75	0.83	4.84	1.38
Marital Status					
married	66.73	71.09	60.74	60.1	66.95
widowed	2.38	1.59	2.5	8.74	2.06
divorced/separated	12.47	10.48	14.82	14.09	13.22
never married	18.43	16.83	21.95	17.07	17.77
Education					
0_11	17.18	12.05	19.91	27.54	21.56
12	39.05	38.5	36.07	42.77	42.63
13_15	23.52	24.43	25.91	19.14	20.05
16+	20.25	25.02	18.11	10.55	15.76
Mean Family Size	3.2	3.2	3.1	3.3	3.4
Mean Age	28.4	28.4	28	29.5	28.3
Employed	0.84	0.84	0.82	0.79	0.87
Labor Income					
Mean	24816	27035	23660	18945	23178
Standard Deviation	24958	29116	23259	22845	16381
Median	21166	22409	21166	15598	21389
Sibling Correlations in 1986	N for health 4511	1692	1178	292	1349
Good health	0.19	0.23	0.19	-0.07	0.12
test all=both in 07 p-value		0.02			
Labor Income	0.22	0.23	0.25	0.1	0.14
test all=both in 07 p-value		0.83			
Education	0.43	0.44	0.45	0.18	0.35
test all=both in 07 p-value		0.16			

Notes: PSID Sample of Adults 18-34 in 1986 who were sample children 0-16 in 1968. Each observation represents a sibling pair. Weighted by 1968 individual weight of older sib. For correlations: goodhealth is good/excellent health on 5 point scale; Labor Income is average over age 25-34; Education is years of education at age 24.

Table 5B: 1986 characteristics by attrition status for Male Sib Pairs

Variable	All	Status in 2007 both in 2007	either out, NR	either out death	either out SEO cut
Charateristics of Child					
Sample	1404	451	387	122	444
Health					
excellent	36.6	39.27	33.47	25.81	38.65
very good	36.97	37.97	44.16	29.14	30
good	20.05	18.61	16.85	24.84	24.72
fair	4.65	3.82	4.38	7.82	5.47
poor	1.73	0.34	1.14	12.38	1.15
Marital Status	70.2	75.59	64.91	63.34	67.58
married	2.1	1.32	1.05	10.08	1.77
widowed	12.04	10.94	12.18	11.25	14.51
divorced/separated	15.67	12.15	21.87	15.33	16.14
never married					
Education					
0_11	18.1	12.45	16.32	30.43	27.02
12	39.11	38.94	34.02	44.92	42.72
13_15	22.42	23.08	28.34	17.58	16.52
16+	20.37	25.52	21.32	7.08	13.74
Mean Family Size	3.26	3.26	3.16	3.42	3.34
Mean Age	28.59	28.79	27.98	29.96	28.31
Employed	0.8915	0.9014	0.893	0.8109	0.9008
Labor Income					
Mean	30422	34742	28551	24072	25662
Standard Deviation	25728	30415	18330	32846	15914
Median	26351	29879	25397	16135	23694
Sibling Correlations 1986	N for health				
	1023	370	242	89	322
Good health	.23*** (.022)	.25***	.12**	-0.08	.18***
test all=in07 pvalue		0.51			
Labor Income	.38***(.016)	.41***	.32***	.18**	.3***
test all=in07 pvalue		0.27			
Education	.43***(.025)	.48***	.45***	-0.01	.26***
test all=in07 pvalue		0.30			

Notes: PSID Sample of Adults 18-34 in 1986 who were sample children 0-16 in 1968. Each observation represents a sibling pair. Weighted by 1968 individual weight of older sib. For correlations: goodhealth is good/excellent health on 5 point scale; Labor Income is average over age 25-34; Education is years of education at age 24.

On correlations: Robust standard errors in parentheses. Asterisks indicate significantly different from zero (**=.05, ***=.01)

Table 5C: 1986 characteristics by attrition status for Female Sib Pairs

Variable	All		Status in 2007 both in 2007	either out, NR	either out death	either out SEO cut
Charateristics of Child						
Sample		1542	615	447	69	411
Health						
excellent		26.76	29.88	19.9	11.92	31.77
very good		35.4	35.85	41.73	39.39	25.53
good		30.74	27.2	31.58	43.67	34.77
fair		6.45	6	6.57	5.02	7.48
poor		0.65	1.07	0.22	0	0.45
Marital Status						
married		64.26	67.32	56.54	61.61	68.54
widowed		1.65	1.44	2.46	0	1.29
divorced/separated		13.12	9.1	19.67	18.88	11.84
never married		20.98	22.14	21.33	19.51	18.33
Education						
0_11		16.61	13.76	19.99	24.38	16.81
12		40.25	37.56	39.6	40.96	46.49
13_15		23.8	25.26	24.99	15.67	20.63
16+		19.34	23.43	15.42	18.99	16.08
Mean Family Size		3.22	3.27	3.02	3.22	3.38
Mean Age		28.18	27.97	28.28	29.04	28.34
Employed		0.7888	0.7738	0.7469	0.8136	0.8711
Labor Income						
Mean		17792	17232	16921	17891	19806
Standard Deviation		13267	12348	11871	16120	15577
Median		14940	14940	14940	11952	16434
Sibling Correlations 1986						
		1330				
Good health		.32*** (.020)	.36***	.26***	.21*	.30***
test all=in07 pvalue			0.15			
Labor Income		.26*** (.021)	.23***	.26***	.32***	.31***
test all=in07 pvalue			0.31			
Education		.43*** (.021)	.46***	.40***	.21*	.43***
test all=in07 pvalue			0..49			

Notes: PSID Sample of Adults 18-34 in 1986 who were sample children 0-16 in 1968. Each observation represents a sibling pair. Weighted by 1968 individual weight of older sib. For correlations: goodhealth is good/excellent health on 5 point scale; Labor Income is average over age 25-34; Education is years of education at age 24.

On correlations: Robust standard error in parentheses. Asterisks indicate significantly different from zero (**=.05, ***=.01)

Table A1

Characteristics of Family Heads of Children Aged 0-16

	PSID 1968	NHIS 1969
Observations on Heads	2774	18436
# of Children	8104	43380
1) Race of Head (%)		
White	83.7	88.7
Black	12.9	10.4
Other	3.4	0.9
2) Education of Head (Years Completed)		
0 to 11	42.9	36.9
12	31.1	35.9
13-15	12.5	12.0
16 or more	13.5	15.3
3) Marital status of Head		
Married	86.8	90.0
Widowed	3.9	2.6
Widow/Divorced/Separated	7.7	6.4
Never Married	1.6	1.1
4) Employed		
Has Job	92.2	90.1
Unemployed/Not in Labor Force	7.8	9.9
5) Family Size		
2	3.0	2.6
3	23.6	23.4
4	26.3	30.3
5 or more	47.0	43.7
6) Number of Children in Family		
1	31.7	33.4
2	28.7	30.9
3	19.4	18.6
4 or more	20.2	17.1
7) Age of Child		
0-5 Years	24.3	24.0
6-12 Years	31.00	35.1
13-18 Years	44.7	40.9

Notes: NHIS sample of family heads , aged 18 or more. PSID heads have children age 0-16 in 1968. PSID original sample members only. Weighted by individual weight.

Table A2

Comparison of PSID and NHIS
Characteristics of Heads/Wives for cohort aged 0-16 in 1968

Study **Year**	**PSID** **1986**	**NHIS** **1986**	**PSID** **1999**	**NHIS** **1999**	**PSID** **2007**	**NHIS** **2006**
Sample Size	4719.00	11795.00	2739.00	21430.00	2605.00	15238.00
Age	18-34	18-34	31-47	31-47	39-55	39-55
1) Race (%)						
White	83.21	84.50	81.00	81.88	82.42	82.36
Black	14.85	10.37	14.94	11.15	15.17	11.57
Other	1.94	5.12	4.06	6.97	2.41	6.07
2) Hispanic (%)	4.43	8.47	1.85	10.22	3.25	11.11
1) Family Size (by head of family)						
1	34.38	33.79	24.56	19.08	27.91	22.71
2	22.57	21.28	20.17	18.69	26.91	26.87
3	18.77	18.86	16.95	20.00	19.48	18.68
4	15.77	16.83	24.10	24.94	16.10	18.78
5 or more	8.51	9.23	14.22	17.30	9.61	12.95
4) Education (Years Completed)						
0 to 11	17.26	14.12	8.95	9.11	12.83	9.76
12	40.91	40.18	38.36	32.31	35.22	31.45
13-15	24.58	25.04	25.12	30.04	24.69	28.40
16 or more	17.25	20.65	27.58	28.54	27.26	30.39
5a) Marital Status Heads & Wives						
Married	64.36	71.05	71.23	76.77	70.36	75.16
Widowed	0.47	0.20	0.78	0.85	1.94	1.22
Divorced/Separated	11.09	7.83	15.30	12.63	16.90	15.26
Never Married	24.07	20.92	12.69	9.75	10.81	8.36
5b) Marital Status All						
Married	67.25	54.74	69.70	70.54	69.26	70.00
Widowed	3.03	0.18	2.03	0.97	2.69	1.42
Divorced/Seperated	11.70	7.48	15.82	14.69	17.23	17.45
Never Married	18.02	37.59	12.45	13.79	10.82	11.17
6) Employment Status						
Has Job	70.83	75.51	69.11	80.79	60.23	78.80
Unemployed or Layoff	8.28	6.47	9.33	4.78	6.42	5.44
Not in Labor Force	20.89	18.02	21.56	14.43	33.34	15.76
7) Income Group (Nominal $)						
< 20,000	31.47	43.31	10.69	12.71	9.86	11.56
≥ 20,000	68.53	56.69	89.31	87.29	90.14	88.44

Notes: NHIS sample of heads and wives of indicated age. PSID sample of heads/wives who were children age 0-16 in 1968 and who have aged to indicated age. PSID includes original sample members only. Weighted by individual weight in each year (final weight in PSID, perweight in NHIS)

Table B1

Comparison of Self-Reported Health in PSID and NHIS

NHIS	**Survey year**		
Health status	1986	1999	2006
Age	18-34	31-47	38-54
Excellent	44.14	36.80	28.01
Very Good	30.08	33.69	33.59
Good	21.13	22.51	27.32
Fair	3.94	5.47	8.25
Poor	0.72	1.53	2.83

PSID SEO+SRC	**Survey year**		
Health status	1986	1999	2007
Age	18-34	31-47	38-54
Excellent	31.38	28.15	18.90
Very Good	37.49	36.40	36.97
Good	24.62	27.13	30.14
Fair	5.78	6.74	10.21
Poor	0.73	1.58	3.78

PSID-SRC only	**Survey year**		
Health status	1986	1999	2007
Age	18-34	31-47	38-54
Excellent	32.62	29.01	19.64
Very Good	38.67	37.21	37.62
Good	22.82	26.18	29.61
Fair	5.31	6.12	9.45
Poor	0.58	1.48	3.69

Notes: Weighted by individual weight (final weight in PSID, perweight in NHIS). Health of family heads/wives indicated age.

Table B2

Percent in Good/Excellent Health by Demographic Groups

NHIS		**Survey year**	
Race-sex	1986	1999	2006
Age	18-34	31-47	38-54
White male	0.80	0.74	0.64
	6164.00	8963.00	6311.00
White female	0.74	0.72	0.63
	6521.00	9441.00	6707.00
Black male	0.64	0.64	0.52
	1069.00	1375.00	1141.00
Black female	0.56	0.59	0.47
	1499.00	1885.00	1562.00
Other male	0.72	0.65	0.64
	379.00	1032.00	698.00
Other female	0.64	0.60	0.61
	418.00	1155.00	730.00

PSID SEO+SRC		**Survey year**	
Race-sex	1986	1999	2007
Age	18-34	31-47	38-54
White male	0.76	0.69	0.61
	812.00	705.00	675.00
White female	0.68	0.67	0.56
	949.00	722.00	688.00
Black male	0.61	0.49	0.54
	512.00	342.00	349.00
Black female	0.48	0.42	0.38
	749.00	579.00	599.00
Other male	0.67	0.84	0.25
	10.00	22.00	9.00
Other female	0.61	0.45	0.41
	30.00	45.00	27.00

Notes: Weighted by individual finalweight in PSID, perweight in NHIS. Unweighted sample size shown below each percentage. Health of family heads/wives indicated age.

Table B3

Linear Probability Model for Good/Excellent Health:
Pooled NHIS and PSID (SEO+SRC) for cohort aged 0-16 in 1968

	Model 1	Model 2	Model 1 Weighted	Model 2 Weighted
Dependent Variable	goodhealth	goodhealth	goodhealth	goodhealth
psid	-0.103***	-0.060***	-0.053***	-0.049***
	(0.009)	(0.009)	(0.011)	(0.011)
Year 1999	-0.041***	0.065***	-0.037***	0.059***
	(0.005)	(0.007)	(0.005)	(0.007)
Year 2007	-0.136***	0.024**	-0.126***	0.021**
	(0.005)	(0.009)	(0.006)	(0.010)
Psid x 1999	-0.005	-0.033**	-0.006	-0.021
	(0.014)	(0.014)	(0.016)	(0.016)
Psid x 2007	0.032**	0.015	-0.004	-0.004
	(0.014)	(0.014)	(0.017)	(0.017)
Age		-0.007***		-0.007***
		(0.000)		(0.000)
White female		-0.033***		-0.030***
		(0.004)		(0.005)
Black male		-0.134***		-0.138***
		(0.008)		(0.009)
Black female		-0.215***		-0.206***
		(0.007)		(0.008)
Other male		-0.046***		-0.040***
		(0.011)		(0.012)
Other female		-0.108***		-0.101***
		(0.011)		(0.012)
Hispanic male		-0.107***		-0.096***
		(0.008)		(0.009)
Hispanic female		-0.117***		-0.112***
		(0.008)		(0.009)
Constant	0.733***	0.988***	0.742***	0.986***
	(0.003)	(0.011)	(0.004)	(0.013)
Observations	65274	65274	65072	65072
R-squared	0.02	0.04	0.01	0.04

Notes: Robust standard errors in parentheses. Weights are normalized to mean of one in each survey sample separately. Hispanic origin can be of any race. Cohort for both surveys is 18-34 in 1986, 31-47 in 1999, 38-54 in 2007 for PSID and 2006 for NHIS.
* significant at 10%; ** significant at 5%; *** significant at 1%

Table D1

Regression of Poor/Fair Health Indicator on background characteristics, SRC Sample Males

Dependent Var =1 if poor/fair health	All in 1991	All in 2007	Diff p value	Sib sample, in 1991	Sib Sample, in 2007	Diff p value	Sib Sample in 1991, Mother FE	Sib Sample in 2007, Mother FE
Low birth weight (<5.5 lbs)	.025**	.032**	.28	.054**	.076**	.03	0.052	0.074
	(0.012)	(0.016)		(0.025)	(0.034)		(0.037)	(0.055)
Faminc0to16 (in 10,000s)	-.001	-0.001**	.19	0.001	-0.001	.14	0.006	0.01
	(0.000)	(0.000)		(0.001)	(0.001)		(0.007)	(0.015)
Mother education =12	-0.011**	-0.002		-0.011	-0.006			
	(0.005)	(0.005)		(0.008)	(0.010)			
=13/15	-0.015***	-0.007*		-0.018**	-0.016*			
	(0.005)	(0.004)		(0.008)	(0.008)			
>=16	-0.014**	-0.003		-0.019**	-0.005			
	(0.006)	(0.005)		(0.009)	(0.009)			
Age	-0.009*	-0.008		-0.013	-0.006		-0.011	-0.006
	(0.005)	(0.005)		(0.009)	(0.009)		(0.011)	(0.012)
Birth order number	0.002	0.003		0.002	0.004*		0	0.001
	(0.001)	(0.002)		(0.001)	(0.002)		(0.007)	(0.010)
Mom married1968	-0.005	-0.011		0.002	-0.002			
	(0.008)	(0.009)		(0.010)	(0.012)			
Agesq	0.017**	0.014		0.023	0.014		0.021	0.014
	(0.008)	(0.009)		(0.015)	(0.016)		(0.018)	(0.020)
Black	0.006	0.016		0.026	0.042			
	(0.009)	(0.012)		(0.019)	(0.026)			
Hisp1968	-0.002	0.013*		-0.008	0.015			
	(0.005)	(0.007)		(0.015)	(0.018)			
Constant	0.132*	0.12		0.168	0.071		0.112	0.002
	(0.071)	(0.076)		(0.133)	(0.130)		(0.159)	(0.143)
Observations	3799	2906		1988	1420		1988	1420
R-squared	0.01	0.02		0.03	0.05		0.01	0.02
Number of group(mother ids)							160	114
Mother fixed effects	No	No		No	No		Yes	Yes

Notes: Robust standard errors in parentheses. * significant at 10%; ** significant at 5%; *** significant at 1%. PSID SRC sample: Child of head age 0 to 16 in 1968 with known mother. Sample years 1986 to 1991. Diff column is p value for test that coefficient from 1991 sample differs from that in 2007 sample. Incomes deflated to 2001 dollars.

Table D2

Outcome Regressions, SRC Males

A. Fair/Poor health (dependent variable)

	Male, All, SRC	Male in 07, All, SRC	P value for Difference	Male, Sibs, SRC	Male in 07, Sibs, SRC	P value for Difference	Male, Sibs xt, SRC	Male Sibs in 07, Sibs xt, SRC
Low birth wt	0.025**	0.032**	0.28	0.054**	0.076**	0.03	0.052	0.074
	(0.012)	(0.016)		(0.025)	(0.034)		(0.037)	(0.055)
Faminc0to16	-0.001	-0.001**	0.19	0.001	-0.001	0.14	0.006	0.01
	(0.000)	(0.000)		(0.001)	(0.001)		(0.007)	(0.015)
Observations	3799	2906		1988	1420		1988	1420
Mother fixed effects	No	No		No	No		Yes	Yes

B. Individual Education (dependent variable)

	Male, All, SRC	Male in 07, All, SRC	P value for Difference	Male, Sibs, SRC	Male in 07, Sibs, SRC	P value for Difference	Male, Sibs xt, SRC	Male in 07, Sibs xt, SRC
Low birth wt	-0.109	-0.041	0.74	0.892*	0.985*	0.74	0.4	0.221
	(0.330)	(0.397)		(0.472)	(0.524)		(0.395)	(0.590)
Faminc0to16	0.165***	0.202***	0.21	0.157***	0.372***	0.004	0.059	0.099
	(0.029)	(0.037)		(0.047)	(0.062)		(0.119)	(0.188)
Observations	728	564		365	257		365	257
Mother fixed effects	No	No		No	No		Yes	Yes

C. Labor Income (dependent variable)

	Male, All, SRC	Male in 07, All, SRC	P value for Difference	Male, Sibs, SRC	Male in 07, Sibs, SRC	P value for Difference	Male, Sibs xt, SRC	Male in 07, Sibs xt, SRC
Low birth wt	-0.174*	-0.329***	0.01	-0.359**	-0.470***	0.34	-0.374*	-0.486***
	(0.092)	(0.103)		(0.143)	(0.132)		(0.194)	(0.136)
Faminc7to16	0.231***	0.257***	0.61	0.424***	0.505***	0.27	0.23	1.027
	(0.073)	(0.082)		(0.085)	(0.105)		(0.494)	(0.656)
Observations	449	354		237	171		237	171
Mother fixed effects	No	No		No	No		Yes	Yes

Notes: Robust standard errors in parentheses. * significant at 10%; ** significant at 5%; *** significant at 1%. PSID SRC sample: Child of head age 0 to 16 in 1968 with known mother. Sample years 1986 to 1991. Diff column is p value for test that coefficient from 1991 sample differs from that in 2007 sample. Models without mother FE also include age and age squared, black, Hispanic, birth order, mother's education, mother's marital status in 1968. Fixed effect models include age and age squared, and birth order. Income, labor income deflated to 2001 dollars.

Table E1

Probits for Being Sample Respondent in 2007, SRC Males

	All	With Brothers	All	With Brothers	All	With Brothers
Hasbadhealth8691	-0.456	-0.116				
	(0.308)	(0.422)				
Individual Education			0.061**	0.082**		
			(0.027)	(0.037)		
Labor Income					0.233**	0.071
					(0.103)	(0.133)
Low birth weight	-0.329	0.109	-0.265	-0.03	-0.377	-0.023
	(0.228)	(0.363)	(0.230)	(0.366)	(0.275)	(0.411)
Faminc0to16	0.003	-0.001	-0.007	-0.01	0.216	0.391
	(0.022)	(0.028)	(0.021)	(0.028)	(0.176)	(0.266)
Age	0.119	0.029	0.069	-0.082	0.556	0.301
	(0.160)	(0.226)	(0.182)	(0.277)	(0.610)	(0.852)
Sibnum1968	0.02	0.112**	0.041	0.160***	0.077	0.134**
	(0.041)	(0.051)	(0.043)	(0.056)	(0.049)	(0.062)
Mother's Education (<12 omitted)						
=12 years	-0.069	-0.049	-0.119	-0.131	-0.191	-0.129
	(0.128)	(0.168)	(0.129)	(0.175)	(0.174)	(0.232)
=13-15	0.085	0.088	0.152	-0.056	0.2	0.015
	(0.202)	(0.289)	(0.218)	(0.308)	(0.277)	(0.418)
=16 or more	-0.089	-0.359	-0.196	-0.469	-0.417	-0.702*
	(0.232)	(0.297)	(0.232)	(0.306)	(0.288)	(0.394)
Mom married 1968	-0.133	-0.141	-0.088	-0.09	0.127	0.111
	(0.231)	(0.288)	(0.210)	(0.289)	(0.276)	(0.356)
Age squared	-0.002	-0.00026	-0.001	0.001	-0.009	-0.005
	(0.003)	(0.004)	(0.003)	(0.004)	(0.010)	(0.014)
Black	-0.111	-0.107	-0.259	-0.205	-0.035	-0.08
	(0.230)	(0.352)	(0.214)	(0.368)	(0.255)	(0.465)
Hisp 1968	-0.843	-0.382	-0.954*	-0.514	-1.168**	-1.225
	(0.545)	(0.741)	(0.488)	(0.763)	(0.559)	(0.752)
Constant	-1	-0.18	-1.08	0.394	-12.31	-9.042
	(2.464)	(3.529)	(2.846)	(4.377)	(9.090)	(12.802)
Observations	709	380	728	365	449	237

Robust standard errors in parentheses. * significant at 10%; ** significant at 5%; *** significant at 1%
Sample of PSID cohort age 0-16 in 1968 with known mother. Sample of respondents present during 1991.
Variable definitions: hasbadhealth8691 is ever had fair or poor health 1986-91; individual education in 1991, respondent at least age 24; labor income averaged over ages 25-34. Income, labor income deflated to 2001 dollars.

Table E2

Probits for Being in Sample in 2007, SRC Females

	All Persons	With Sisters	All Persons	With Sisters	All Persons	With Sisters
Hasbadhealth8691	-0.14	-0.374				
	(0.306)	(0.363)				
Individual Education			0.042	0.027		
			(0.030)	(0.042)		
Labor Income					0.035	-0.022
					(0.061)	(0.071)
Low birth weight	0.028	0.129	-0.031	-0.001	0.097	0.267
	(0.190)	(0.224)	(0.192)	(0.235)	(0.239)	(0.257)
Faminc0to16	0.027	0.032	0.017	0.023	0.129	0.29
	(0.019)	(0.027)	(0.019)	(0.029)	(0.158)	(0.211)
Age	0.061	-0.095	0.096	-0.092	0.441	0.569
	(0.159)	(0.209)	(0.178)	(0.244)	(0.566)	(0.712)
Sibnum1968	-0.013	0.011	-0.004	0.029	-0.018	0.005
	(0.043)	(0.053)	(0.044)	(0.054)	(0.049)	(0.059)
Mother's Education						
(<12 omitted)						
=12 years	0.047	0.122	-0.009	0.117	-0.193	-0.055
	(0.128)	(0.158)	(0.134)	(0.165)	(0.177)	(0.213)
=13-15	0.673***	0.799***	0.548**	0.731**	0.593*	0.693*
	(0.233)	(0.283)	(0.242)	(0.298)	(0.330)	(0.384)
=16 or more	0.039	0.443	-0.146	0.387	-0.188	0.192
	(0.214)	(0.291)	(0.227)	(0.309)	(0.285)	(0.363)
Mom married 1968	-0.057	-0.454*	-0.003	-0.366	-0.616	dropped
	(0.211)	(0.269)	(0.210)	(0.273)	(0.376)	
Age squared	-0.001	0.001	-0.002	0.001	-0.008	-0.01
	(0.003)	(0.003)	(0.003)	(0.004)	(0.010)	(0.012)
Black	-0.006	0.049	-0.181	0.065	-0.114	0.001
	(0.197)	(0.223)	(0.184)	(0.223)	(0.245)	(0.295)
Hispanic	dropped					
Constant	-0.151	2.283	-1.188	1.762	-6.473	-10.36
	(2.459)	(3.249)	(2.825)	(3.905)	(8.550)	(10.765)
Observations	728	411	712	388	465	262
Pseudo R2	0.02	0.04	0.02	0.04	0.03	0.04

Robust standard errors in parentheses. * significant at 10%; ** significant at 5%; *** significant at 1%
Sample of PSID cohort age 0-16 in 1968 with known mother. Sample of respondents present during 1991.
Hispanic indicator dropped because all SRC Hispanic females leave sample.

Variable definitions: hasbadhealth8691 is ever had fair or poor health 1986-91; individual education in 1991, respondent at least age 24; labor income averaged over ages 25-34. Income, labor income deflated to 2001 dollars.

Appendix Table D3A

Outcome Regressions, SRC Females

A. Poor/Fair Health Indicator (dependent variable)

	Female, All, SRC	Female in 07, All, SRC		Female, Sibs, SRC	Female in 07, Sibs, SRC		Female, Sibs xt, SRC	Female in 07, Sibs xt, SRC
lowbirthwgt	-0.001	-0.004	0.29	-0.001	-0.005**	0.48	-0.002	-0.002
	(0.004)	(0.004)		(0.005)	(0.002)		(0.001)	(0.002)
faminc0to16	0	0	0.18	0	0	0.14	0	0.002
	(0.000)	(0.000)		(0.000)	(0.000)		(0.001)	(0.001)
Observations	4137	3249		2312	1584		2312	1584

B. Individual Education (dependent variable)

	Female, All	Female in 07, All		Female, Sibs	Female in 07, Sibs		Female, Sibs xt	Female in 07, Sibs xt
lowbirthwgt	-0.348**	-0.352	0.98	-0.099	-0.213	0.59	0.615***	0.168
	(0.170)	(0.237)		(0.190)	(0.262)		(0.235)	(0.316)
faminc0to16	0.153***	0.151***	0.86	0.179***	0.188***	0.66	0.108	0.256***
	(0.019)	(0.023)		(0.023)	(0.030)		(0.070)	(0.095)
Observations	1561	934		943	467		943	467

C. Labor Income (dependent variable)

	Female, All, SRC	Female in 07, All, SRC		Female, Sibs, SRC	Female in 07, Sibs, SRC		Female, Sibs xt, SRC	Female in 07, Sibs xt, SRC
lowbirthwgt	0.366**	0.302	0.45	0.469**	0.489***	0.89	0.524**	0.308
	(0.174)	(0.194)		(0.192)	(0.177)		(0.232)	(0.289)
faminc7to16	0.457***	0.445***	0.84	0.503***	0.621***	0.35	-0.628	0.022
	(0.121)	(0.133)		(0.162)	(0.199)		(0.854)	(1.002)
Observations	469	379		276	191		276	191

Appendix Table D3B

Outcome Regressions: SRC + SEO Sample, Males

A. Bad Health sample (dependent variable)

	Male, All	Male in 07, All	Diff p value	Male, Sibs	Male in 07, Sibs	Diff p value	Male, Sibs xt	Male in 07, Sibs xt
lowbirthwgt	0.014*	0.040***	0.004	0.031**	0.095***	0.0003	0.057	0.101
	(0.008)	(0.015)		(0.014)	(0.031)		(0.038)	(0.064)
faminc0to16	0	0	0.55	0.001	0.002	0.46	0.003	0.007
	(0.001)	(0.001)		(0.001)	(0.001)		(0.005)	(0.011)
Observations	7337	3891		4076	1894		4076	1894

B. Individual Education (dependent variable)

	Male, All	Male in 07, All		Male, Sibs	Male in 07, Sibs		Male, Sibs xt	Male in 07, Sibs xt
lowbirthwgt	-0.133	-0.014	0.56	0.098	0.449	0.26	-0.006	0.018
	(0.189)	(0.296)		(0.295)	(0.477)		(0.287)	(0.525)
faminc0to16	0.182***	0.206***	0.41	0.174***	0.337***	0.002	0.074	0.073
	(0.025)	(0.035)		(0.037)	(0.054)		(0.108)	(0.156)
Observations	1451	772		763	345		763	345

C. labor (dependent variable)

	Male, All	Male in 07, All		Male, Sibs	Male in 07, Sibs		Male, Sibs xt	Male in 07, Sibs xt
lowbirthwgt	-0.222**	-0.263***	0.66	-0.241**	-0.397***	0.16	-0.209	-0.372***
	(0.096)	(0.087)		(0.114)	(0.098)		(0.155)	(0.083)
faminc7to16	0.365***	0.311***	0.43	0.508***	0.402***	0.27	0.393	0.227
	(0.061)	(0.082)		(0.080)	(0.093)		(0.388)	(0.504)
Observations	950	493		518	229		518	229

Appendix Table D3C

Outcome Regressions: SRC + SEO Sample, Females

A. Bad Health sample (dependent variable)

	Female, All	Female in 07, All		Female, Sibs	Female in 07, Sibs		Female, Sibs xt	Female in 07, Sibs xt
lowbirthwgt	0	-0.004	0.31	-0.004	-0.007***	0.42	-0.021	-0.001
	(0.004)	(0.003)		(0.004)	(0.002)		(0.014)	(0.001)
faminc0to16	0.001	0	0.12	0.001	0	0.02	0.012	0.003
	(0.000)	(0.000)		(0.000)	(0.000)		(0.008)	(0.002)
Observations	8862	5200		5535	2745		5535	2745

B. Individual Education (dependent variable)

	Female, All	Female in 07, All		Female, Sibs	Female in 07, Sibs		Female, Sibs xt	Female in 07, Sibs xt
lowbirthwgt	-0.348**	-0.352	0.98	-0.099	-0.213	0.59	0.615***	0.168
	(0.170)	(0.237)		(0.190)	(0.262)		(0.235)	(0.316)
faminc0to16	0.153***	0.151***	0.86	0.179***	0.188***	0.66	0.108	0.256***
	(0.019)	(0.023)		(0.023)	(0.030)		(0.070)	(0.095)
Observatioions	1561	934		943	467		943	467

C. Labor (dependent variable)

	Female, All	Female in 07, All		Female, Sibs	Female in 07, Sibs		Female, Sibs xt	Female in 07, Sibs xt
lowbirthwgt	0.027	0.182	0.11	0.083	0.333**	0.08	0.17	0.383*
	(0.127)	(0.156)		(0.146)	(0.134)		(0.216)	(0.202)
faminc7to16	0.476***	0.534***	0.44	0.480***	0.628***	0.24	0.235	1.104
	(0.093)	(0.120)		(0.127)	(0.165)		(0.789)	(1.081)
Observations	1026	637		654	340		654	340