NIH Public Access
Author Manuscript
B E J Econom Anal Policy. Author manuscript; available in PMC 2012 February 23.

Published in final edited form as:
B E J Econom Anal Policy. 2011 ; 11(3): . doi:10.2202/1935-1682.2868.

Attrition in Models of Intergenerational Links Using the PSID with Extensions to Health and to Sibling Models*

John M. Fitzgerald*

Abstract

Selective attrition potentially biases estimation of intergenerational links in health and economic status. This paper documents attrition in the PSID through 2007 for a cohort of children, and investigates attrition bias in intergenerational models predicting adult health, education and earnings, including models based on sibling differences. Although attrition affects unconditional means, the weighted PSID generally maintains its representativeness along key dimensions in comparison to the National Health Interview Survey. Using PSID, sibling correlations in outcomes and father-son correlations in earnings are not significantly affected by attrition. Models of intergenerational links with covariates yield more mixed results with females showing few robust impacts of attrition and males showing potential attrition bias for education and earnings outcomes. For adult health outcomes conditional on child background, neither gender shows significant impacts of attrition for the age ranges and models considered here. Sibling models do not produce robustly higher attrition impacts than individual models.

Keywords

attrition; PSID; health and SES; intergenerational correlations; intergenerational models

1 Introduction

The Panel Study of Income Dynamics (PSID) has become a premier data set for investigating intergenerational transmission of social and economic status. Because the PSID has followed families and descendants since 1968, it provides a long time frame over which analysts can observe family backgrounds for children as well as adult outcomes for those children. Many influential studies have found significant intergenerational linkages in income and status using data from the PSID (Solon, 1992; Solon, 1991; Behrman, 1990). Links between poor child health and low parental income and later adult income and health outcomes are also well established (Currie, 2009). Understanding the nature of these intergenerational correlations in health and income is important in order to formulate appropriate policy responses. Estimation of the linkages requires strong data that follow individuals from childhood to adulthood and permit good controls for family background and neighborhood environment.

In extended longitudinal studies such as the PSID, attrition by respondents potentially damages the representativeness of the survey over time. Attrition is a particular concern in

*Funding from the Small Grants Program for Research Using PSID Data is gratefully acknowledged. It bears no responsibility for errors or conclusions. Thanks to participants at the conference on “SES and Health across Generations and over the Life Course,” at the University of Michigan, Ann Arbor, September 2010. Thanks to Matthew Delaney and William Jacob for excellent research assistance.

Copyright © 2011 Berkeley Electronic Press. All rights reserved.

*Bowdoin College, jfitzger@bowdoin.edu.

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

intergenerational models because of the long time frame needed to observe both child background and adult outcomes. For example, past work has shown that attrition is higher among the less educated and those with lower incomes, both commonly studied adult outcomes. In addition, attrition is also higher for those in ill health (Reynolds, Frank et al., 2005). The primary question for this paper is the extent to which attrition has biased estimates of intergenerational models that relate child background to adult outcomes using the PSID.

This paper will update some previous work on attrition for children who age to adulthood during the PSID, a sample of primary interest for intergenerational studies (Fitzgerald, Gottschalk, Moffitt, 1998b). In addition, the paper extends the literature by looking at the impact of attrition on health outcomes in adulthood and how these adult outcomes depend on family background. To condition on family background, a number of studies (cited below) use family or mother fixed effects. These models require the observation of at least two siblings as adults which adds to the scope of the attrition problem. A further contribution of this paper is to investigate the effects of attrition on sibling pairs and in mother fixed effect models.

The paper begins by documenting attrition rates in the PSID and how attrition has affected unconditional means for characteristics over time for the cohort of children aged 0 to 16 in at the beginning of the PSID. This group is the first cohort of child to age to adulthood during the panel and is the primary focus of this paper. To gauge the potential for bias by attrition, the paper next compares health and SES measures in various years for this cohort to nationally representative cross-sectional data from the National Health Interview Study. Finally the paper focuses on intergenerational models of how family background affects adult outcomes. It tests the degree to which attrition in the PSID might bias estimation of these intergenerational linkages. Before turning to the empirical analysis, the next section provides background on attrition studies and intergenerational models.

2 Background

Attrition in the PSID

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 parameters based on the surviving sample. As will be documented below, attrition in the PSID has been substantial. Of all children aged 0 to 16 in the first year of the panel in 1968, less than one-third remain in 2007. Of those in the nationally representative SRC sample component (explained below) about half remain. Attrition has long been a concern, and many studies have documented it and discussed methods for dealing with it (Becketti, Gould et al., 1988; Fitzgerald, Gottschalk, and Moffitt, 1998a; Lillard and Panis, 1998; Ziliak and Kniesner, 1998). Past work based on the PSID has shown that attrition is higher among minorities, the less educated, those with lower incomes, and who move. But this alone does not indicate a problematic bias for two reasons. Although substantial attrition affects the unconditional means of many social and economic variables, the PSID survey weights are remarkably good at preserving sample representativeness. Furthermore, models of earnings, education, or welfare recipiency that condition on a rich set of demographic covariates tend to show little impact of attrition on parameter estimates (Fitzgerald, Gottschalk, Moffit, 1998a,b). But these studies have not directly investigated health.

Attrition and Health

Several studies have commented on the relation between attrition and health. 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; Alderman, Hoddinott, Kinsey, 2006).

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Thus, through time, one might expect that the PSID respondents become healthier than a representative cross-sectional sample. (A comparison is made in Part 7 below.) Based on the PSID, Halliday and Kimmitt (2008) confirm that less healthy people are more likely to drop out, but also find that less healthy people are less mobile. Less mobility may make them easier to track over time. Meer, Miller et al. (2003) use the PSID and find a weak link between wealth and health. They test for attrition impacts by estimating the relation separately for those who later attrite and those who remain, and conclude that attrition is not driving their results. Johnson and Schoeni (2007) also use the PSID and develop models for a sample of children relating birth weight, a potential indicator of child health, to later adult outcomes. They report that birth weight was not predictive of which children remained in the sample to adulthood, and conclude that attrition was not selective by birth weight.

Intergenerational Models

Intergenerational models relate parental background and income to adult outcomes for children followed to adulthood. A large literature investigates parent child correlations in earnings, education, welfare usage, and more generally economic status. A fairly recent literature demonstrates that family background and income affect health as well as other status variables[1].

To the extent that attrition selects on income, earnings, wealth, or health, as the children age, we may not retain representative observations on adult outcomes. This could compromise the intergenerational analysis, but need not as explained by Fitzgerald, Gottschalk, Moffitt, 1998b, hereafter "FGM." To see this, consider their intergenerational model that relates adult outcomes in year t to child background during childhood year s as follows:

$$y_{it}=\beta_0+\beta_1 xc_{it}+\beta_2 xp_{is}+\varepsilon_{it} \quad (1)$$

where y_{it} is an outcome for child i as an adult in year t,

xc_{it} stands for characteristics of the adult child in year t, and

xp_{is} stand for characteristics of the parents of child i in childhood year s.

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

$$A*_{it}=Z_{fit}\delta+v_{fit} \quad \text{and}$$
$$A_{it}=1 \text{ for non}-\text{response in year t when } A*_{it}>0 \text{ and}$$
$$=0 \text{ for observed response when } A*_{ft}<=0.$$

The vector Z could include xc and xp as well as additional identifying variables not included in the structural model (loosely called instruments). As discussed in FGMa, attrition biases estimates of the structural relation when $E(\varepsilon|xc, xp) \neq 0$. For example, bias results if attrition selects out those with low earnings *given their parent's earnings*. Thus the extent of bias in models of this type depends on the conditioning variables and is model specific.

FGMb investigate the role of attrition in intergenerational correlations in labor income, education, and welfare recipiency using the PSID, and find that the weighted second generation is generally representative. However, models of intergenerational correlations in

[1]Health during childhood is a significant determinant of adult outcomes (Currie, 2009; Smith, 2009; Johnson and Schoeni, 2007). 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).

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

variables like education yield more mixed results and may be sensitive to attrition. This paper will proceed along similar lines, developed below, with additional consideration of health and of mother fixed effect models.

Intergenerational Models with Family Effects

Determining the causal path from parental Social and Economic Status (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 both poor child health and low family income, or poor child health could lead to reduced maternal employment and low income (Currie, 2009).

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, 2009). Family or mother fixed effects and models that eliminate family effects by use of differences across siblings will later be referred to as sibling models. Sibling models difference out confounding common permanent income and parent effects that might cause spurious correlations. Sibling models obviously are not of value when the parental characteristic of interest does not vary across siblings. But in circumstances that vary across siblings (e.g. family income when child is a specified age), sibling models potentially allow identification of impacts[2].

Attrition causes two difficulties for sibling models. The first is simply that sibling models require that more than one sibling be retained in the sample. Thus sibling pairs have a lower probability of retention than one individual and this intensifies attrition. Second, sibling differences in outcomes and covariates may be affected by attrition. For example, sibling pairs that survive in the panel to adulthood may be more similar if attrition selects out siblings that have very high or very low income (or health) so that those pairs are lost. This would make it more difficult to get useful variation across siblings.

A minor extension of the model above helps clarify some aspects. Consider a generic model that relates adult outcomes in year t to child background during childhood year s as follows:

$$y_{fit}=\beta_0+\beta_1 xc_{fit}+\beta_2 xc_{fis}+\beta_3 xp_{fis}+\alpha_f+\varepsilon_{fit} \quad (2)$$

with notation as before except subscript fit referring to child i in family f as an adult in year t and subscript fis referring to child i in family f in childhood year s, and α_f standing for family fixed effects.

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 that includes a family fixed effect η_f

[2]Several problems complicate this identification. First, differencing can exacerbate measurement error (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 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 siblings during childhood may be due to endogenous factors related to their own or parents health (Smith, 2007; Johnson and Schoeni, 2007).

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

$$A*_{fit}=Z_{fit}\delta+\eta_f+\nu_{fit} \quad \text{and}$$
$$A_{fit}=1 \text{ for non-response in year t when } A*_{fit}>0 \text{ and}$$
$$=0 \text{ for observed response when } A*_{fit}<=0.$$

Estimation of a fixed effect model requires adult responses for two or more siblings and thus requires the joint event $A_{fit}=0$ and $A_{fjt}=0$ for at least two siblings i and j.

Under restrictive assumptions, the selection problem can be avoided in this model. 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 $\Delta\varepsilon$ 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 sibling and this will result in a selection bias after differencing. This more general case is the one pursued in this paper.

3 Approach of this study

This paper follows the line of Fitzgerald, Gottschalk, Moffitt (1998a, 1998b), and focuses on attrition as a problem of "selection on observables"[3]. This assumes that selection is based on a vector of variables Z that affects the attrition probability, and is correlated with the ε but is not part of the "structural" model. In general, these Zs can be lagged values of the dependent variables (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)[4]. Construction of weights or other solutions is left for future work while this paper focuses on documenting attrition and its impacts.

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 for attrition bias, we investigate selection on both unobservables and observables by comparing the distribution of selected variables from the PSID to those from

[3]The alternative approach of "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). See discussion of the distinction in FGMa.

[4]That is, the analyst estimates the probability $P(A_{ff}=0|z_{fi})=p_{fi}$ and weights by $1/p_{fi}$. In a sibling difference model, however, the individual weights may be inadequate because sample selection requires that both siblings survive to the adult outcome period. Thus the proper weight requires that we estimate the joint probability $P(A_{ff}=0, A_{fj}=0|z_{fi}, z_{fj})$. These joint weights will differ from the product of the individual 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. By way of comparison, these would not be suitable "instruments" in a selection on unobservables framework where the Zs cannot correlate with $\Delta\varepsilon$.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

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. We find that the data sets are similar in some 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 variable). 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, sample retention probits are estimated 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 income 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.

4 Data Issues

Measurement of intergenerational links requires 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, ages to 39–55 by 2007. It will reach ages 18–34 in 1986, the first year when health measures are available for all family members. For the intergenerational models, the sample is persons listed as child of the family head in 1968. 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 uses commonly employed adult outcome measures: adult self-assessed general health, earnings, and education along the lines of Smith (2009), Haas (2006), Johnson and Schoeni (2007). Specifically, in the adult outcome models, we measure adult 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 (labor income excluding farm and business income averaged over the ages 25 to 34). The averaging over time helps limit measurement error (Solon, 1991). Because parent child correlation in earnings is the subject of a large literature, this paper also presents alternate specifications of father-son earnings correlations (different age ranges, etc.) for comparison to those in the literature. Incomes and labor incomes are deflated to 2001 dollars using the GDP consumption deflator, except when indicated as nominal.

For the primary models, background characteristics include child birth weight, parental average income during childhood, parental education, parental age, sibling ages and birth order effects. A number of studies have focused on the role of birth weight and parental

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

income (Currie, 2009)[5]. Although some of these studies suggest that interactions are important, this paper will use a specification that includes family income, mother education, and child birth weight without interactions as an easier to interpret model.

Health data have 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[6]. In 1999, a retrospective health question about childhood health was asked (general health when respondent was less than 17). Smith (2009) and Haas (2006) both use the 1999 PSID retrospective health module 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 in 1985 as a binary variable for low birth weight (less than 5.5 pounds). For children born after 1985, a continuous measure of birth weight is collected in the 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). 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[7]. For both of these reasons, results from the SRC subsample are emphasized in this paper, although some tables include the combined SEO and SRC samples as 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.

The Data Appendix summarizes sample and outcome variable definitions for the various tables in the paper.

5 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

[5]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 adult health. They also find that maternal birth weight and income at birth interact to affect child's birth weight.

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

[7]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).

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

samples. There are 8104 children in the 1968 cohort, with 7527 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 includes deliberate sample cuts mentioned above. This simply tabulates the proportion responding; consequently, 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 siblings age 16 or less present in 1968 ("with siblings"). 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 somewhat 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 sibling in that year. That is, the prior column shows response rates for children who initially had siblings in 1968, regardless of these siblings' 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 explicit 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 revise the base to account for mortality for those who have left the sample. This correction is not pursued in this paper,[8] but the second panel of Table 1 shows response rates for a sample that excludes 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 siblings remaining in the sample. In the top panel children with siblings appear to have higher response rates than those with no siblings, 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

[8]See FGM for an example when this mortality correction is applied is an earlier PSID sample.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

to matter little. The extent to which characteristics of the remaining respondents differ from those who attrite is developed further in the next section.

6 Characteristics by Attrition Status

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[9]. 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 children 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. The PSID provides sample weights that were recomputed post 1997 to compensate for this cut (Heeringa and Connor, 1999).

Table 4 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-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 aged 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. Comparing the full sample to those remaining is more relevant in assessing whether the remaining sample is representative. Testing reveals that the health distribution of the complete sample and those

[9]Tables available from author.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

who remain in 2007 are statistically different[10],[11], but substantively they appear quite similar. Thus attrition appears to be mildly 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 in health between those who are observed to remain in the panel to 2007 compared to those that leave, even excluding those known to die. A hypothesis test rejects that the health distributions of the complete sample and those that remain in 2007 are the same[12]. Thus attrition is 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[13].

Attrition by Sibling Pairs

Sibling pair data can be used to eliminate family fixed effects and to calculate sibling correlations. The next set of tables display results for sibling pairs. The sample consists of one observation for each unique sibling pair, hence 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[14]. The bottom part of the table show correlations between the siblings 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 siblings only and female siblings only, respectively. All three tables show that siblings 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[15]. For females, we observe less of a difference in labor income. Sample retention appears to favor the more advantaged, and healthier, but the differences in health are not large in size.

[10]The difficulty in testing is that the sample in 2007 is a subset of the 1986 sample, so the samples overlap. The first test is simply a Pearson Chi2 for hypothesis that the distribution of health is the same in the complete sample in 1986 and those remaining in 2007. The relevant Chi2 would be the contribution of those remaining in plus the contribution from the complete sample. Since the complete sample is used to derive the marginals for the Chi2 test, it contributes nothing to the sum. So the contributions to the Chi2 are based on the complete versus those retained in 2007 samples (not the typical in versus out chi2). The Chi2(4) statistic is 9.97 so we reject that the distributions are the same at the .05 level.

[11]The second method is to estimate an ordered logit for each sample (complete and in 2007) in a model with only cutoffs and no covariates. We then test the hypothesis that the cut points of the ordered logit are the same in the two samples. This is a test across overlapping samples, and thus Stata's suest command is used to get the Wald statistic. The test uses 1968 sample weights and adjusts for sample design. The adjusted F(4,29) is 2.76 which is significant at the .05 level.

[12]By same methods as footnote 10, the Chi2(4) for fathers is 76.8 and for mothers is 76. Both are significant at the 1 percent level.

[13]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 are available from the author that show attrition rates by parents and children.

[14]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 known to have died, then assigning remainder to non-response to those with neither SEO cut or death occurring.

[15]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.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

The bottom panels of Tables 5B and 5C display sibling correlations based on 1986 data. Sibling correlations measure the importance of common sibling background and hence reflect intergenerational links (Solon et al., 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[16], and for education the correlation of .43 for the complete sample rises to .48 for the subsample that remain in 2007. Thus the correlations rise slightly in the retained subsample, but the difference between the full sample and the sample with both siblings responding in 2007 are not statistically significant[17]. (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[18].

This section has 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.

7 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 close. But that study did not consider health measures. Andreski et al. (2007) compare PSID and NHIS data on dimensions including health; this section reports on and extends their work.

The NHIS is chosen for the comparison because it is a large repeated cross-section survey with health and demographic information. The repeated cross-sections from the NHIS are

[16]Solon et al. (1991) estimate the correlation in permanent earnings for brothers using the PSID based on a variance components model. They obtain value of .34 based OLS estimation and a value of .45 in a model that allows for serial correlation. They note these correlations exceed correlations based on single year earnings. The male earnings correlation in Table 5B (.38) is based on muli-year averages for each brother at age 25–34 and resembles Solon, et al’s estimate of the correlation in permanent earnings. Solon et al. do not report a correlation for women’s labor income, but they find a correlation in permanent sister’s income of .28. This is similar to my Table 5 C value of .26 for correlation in sister’s labor income.

[17]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 are estimated as a standardized beta from a regression coefficient of one sibling’s outcome on the other sib’s outcome. Testing the difference in coefficients across the two samples is complicated because the samples overlap and are not independent. The test uses Stata’s suest capability to calculate the combined robust covariance matrix allowing for non-independent samples and performs a Wald test.

[18]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 pooled gender sample.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

assumed to show the distributions of characteristics for a sample not subject to attrition[19] in the various years. One non-comparability is that the cross-sectional cohorts in NHIS allow for immigration, but by construction the cohort in PSID does not. That is, the PSID weights restore the 1968 demographic distributions. Thus, differences in sample demographic composition will affect some comparisons, as noted[20].

This section first compares the characteristics of the children aged 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 specified cohorts from the three NHIS survey years are then compared to the same cohorts 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 has released data through 2007 as of this writing. 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 6 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 to non-comparable definitions[21]. 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 exception of education, the samples are roughly similar.

In Table 7, the characteristics of the child cohort are shown as it ages into adulthood. Changes in race percentage reflect both variation in definitions over time and changes in sample composition since NHIS allows immigration and PSID does not. 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[22]. Lastly, family size is similar in 1986, but over time the NHIS shows larger family sizes than PSID. Again, this could reflect the inclusion of immigrant families in the NHIS.

[19]This ignores that initial period non-interviews occur, with a frequency that may differ across surveys.
[20]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 because we do not observe characteristics during childhood together with adult outcomes for these groups.
[21]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.
[22]Nominal income is compared because the NHIS uses nominal income brackets.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Overall, it appears that the weighted PSID sample maintains its representativeness over time along several key characteristics. 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.

Comparison of Health Measures in PSID and NHIS

Andreski et al. (2007) compare responses on health-related questions for the PSID and 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 difference. They note that the 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 8 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 8 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 aged 0–16 in 1968, 18–34 in 1986, 31–47 in 1999, and 38–55 in 2007 for PSID and 2006 for NHIS. Table 8 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 the percentage of those in good or excellent health (the age gradient) is about the same, a point developed further below. We next explore factors that might reconcile the surveys.

Table 9 shows general health by demographic group. The demographic composition of the PSID is fixed in 1968 and these tables exclude new sample members. To see if varying the demographic composition of the surveys might explain the lower health in PSID, Table 9 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 10 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

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

year with the PSID indicator 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 two condition 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. Conditioning on demographic variables reduces the PSID indicator (to −.06 from −.10) 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 significant difference in the trend of health for the PSID and NHIS in the model without demographic indicators. With demographic indicators there is little difference in the trend at 2007 but a small difference at 1999, barely significant at the 10 percent level.

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 once the initial health difference is removed.

Overall, the analysis suggests that although PSID may have 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 as they age, 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.

8 Adult Outcome Models

A key difficulty in the intergenerational models of interest in this paper is that we do not observe adult outcomes until the child cohort has aged to adulthood in the panel. And we do not observe baseline health variables for all respondents until 1986 or birth weight until 1985. The sample thus has been subjected to significant attrition before we observe the adult outcomes of interest. As discussed in FGMb, in order to proceed with tests for attrition with this type of intergenerational data we need a strong assumption. Consider a year r part way through the panel (r<t) as a baseline year (e.g. 1991). 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. So 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 who have reached adulthood. For our purposes we choose 1991 as the baseline because we then have access to child birth weight as well as parental background for the pre-r period, and we can observe adult outcome variables including health, earnings, and

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

education for the child cohort in young adulthood. To get better measures of outcomes, the method takes advantage of multiple years of outcomes and uses the period 1986 to 1991 for health, years of education for respondents in 1991 when they are at least age 24, and earnings averaged over the ages 25–34 for respondents in 1991.

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

$$y_{fir}=\beta_0+\beta_1 xc_{fir}+\beta_2 xc_{fis}+\beta_3 xp_{fis}+\alpha_f+\varepsilon_{fir} \quad \text{for } A_{fir}=0 \text{ and } A_{fjr}=0. \tag{3}$$

The same year r relation is then estimated for the respondents who have survived to time t > r (with $A_{fit}=0$ 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 begin with an analysis of the retention of individuals. We then turn to an analysis of retention of individuals together with siblings and with sibling pairs.

Outcome Regressions

Father Son Earnings Correlations—Before turning to more complicated models, this section begins by exploring correlations in father's and son's earnings, a common measure of intergenerational links with a large literature[23]. The approach below uses multiple year averages of both father and son earnings to reduce measurement error (Solon, 1992). Furthermore, results are shown using two different age ranges of the fathers to help control for "life cycle" bias due to systematic measurement error over age profiles (Haider and Solon, 2006). The method below most closely resembles Gouskova, Chiteji, Stafford (2010) who use the PSID and report elasticities in 5 year averages that accounts for life-cycle bias by matching father's and son's ages for 10 year cohorts.

Table 11 shows earnings elasticities computed based on log of average earnings for sons age 25–34. To insure that we have at least five years of earnings data for the sons, the sample is restricted to sons who turn age 25 by 1986 and we calculate the average for sons who are in sample in 1991 (the base period). For fathers we show averages at two different age ranges, as well as an elasticity based on family income when the son was a child. The estimates are fairly consistent with those found in the literature, but models are not exact replications because of sample definition differences[24], [25].

As for attrition, the main point of this section is that when we compute the same earnings elasticities restricting the sample to those who remain in the panel in 2007, the correlations do not change. That is, Table 11 shows that those in the selected sample of non-attritors in 2007 produce the same elasticities as the larger "complete" 1991 sample. Correlation sizes

[23]Thanks to an anonymous referee for suggesting a section on parent child earnings elasticities.

[24]Solon (1992) obtains an earnings elasticity of .41 based on 5 year averages using PSID. Gouskova, Chiteji, Stafford (2010) also use PSID and report an elasticity of .29 for samples drawn when father is age 25–34 and son is age 25–34 based on correlation in a 5-year average of log earnings. I obtain a larger value of .47, but our samples are not the same because I include only observations where the son was age 25 by 1986 (born 1952–1961) whereas they include sons born 1956–1979 which produces a larger sample. I also condition on age and age squared for father and son within the 10-year cohort, and calculate the average of the logs based on all years available for each cohort instead of randomly selecting 5 years. For samples of older fathers age 35–44 and sons age 35–44 they get a higher estimate of .41. They note that their estimates are lower when the father's cohort is at an older age than the son's which is consistent with that reported here. They also note that using the log of average income, which I use throughout the rest of the paper, results in slightly lower elasticities, also consistent with the table here.

[25]The estimates of impact of family income on son's earnings presented in the next section are generally lower than those here because later models condition on other background variables besides father and son age.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

are similar and a hypothesis test shows no significant difference between the correlation for the full 1991 sample and selected 2007 sample. This is one indication that attrition is not biasing intergenerational correlations in earnings, at least for attrition occurring late in the panel[26],[27].

Health, Education, and Labor Incomes—A number of recent articles based on PSID data estimate intergenerational models using family income when a child as a predictor of later adult outcomes. Smith (2009) relates parental income when the child was aged 0 to 16, parental education, and other background characteristics to adult health, education, earnings, income, and wealth as an adult. Johnson and Schoeni (2007) and Haas (2006) also relate birth weight and measures of family income and SES when a child to adult health, earnings, education and other outcomes measures. This section uses regression specifications in that style relating parental education, income and child birth weight to adult outcomes and then asks whether selection by attrition affects the coefficients.

Three outcome variables are used in the structural models: 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 aged 0 to 16, mother's education, and child birth weight[28]. 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[29]. In Table 12, 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 log of average family income when the child was aged 0 to 16 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. The effects for mother's education are reduced. The test for attrition bias tests whether the coefficients in the "full sample" model of column one differ from those of the selected 2007 sample[30]. Neither the coefficient for birth weight or for family income is found to be significantly different across the two samples (p-value shown in table).

---

[26]This finding is consistent with FGM based on PSID data up through 1989.

[27]Fixed effect estimates for these correlations are not reported because the parental variables would be the same across siblings (i.e., earnings at specific ages of the father).

[28]Although average family income uses a different number of observations for children at different ages, it measures childhood environment and is the family income construct used in Smith (2009). Alternative averaging methods might change results somewhat, but what matters for attrition comparisons is that the computation is done in the same way for the complete and selected sample.

[29]The multiple year window is used by Johnson and Schoeni (2007) to study health outcomes. 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.

[30]The method uses Stata's suest capability to calculate a combined covariance matrix with overlapping samples and then uses a Wald test as described in footnote 17.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Results for education and for labor income for the Male SRC sample are shown in next columns. In the education regression, high average family income during childhood and mother's education have positive effects on adult educational attainment. Attrition effects are insignificant: birth weight and log family income coefficients are not significantly different between the complete and selected sample. Turning to the model for SRC 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 its impact becomes larger in absolute value and statistically significantly different when the selected 2007 sample is used. Family income also has a significant positive effect on adult labor income, but there is not a statistically non-zero difference between the complete and retained 2007 sample[31].

Results for SRC women are presented in Panel B of Table 12. Neither birth weight nor family income are significant predictors of adult health for women. For education and labor outcome variables, family income coefficients are stable and indicate that higher family income while a child increases education and earnings even after conditioning on other variables. The results for birth weight are somewhat unstable and have counterintuitive signs for health and labor outcomes. This suggests caution in interpreting the birth weight results for females. The test for attrition bias (whether the 1991 and 2007 samples produce statistically different coefficients) does not reveal significant differences in coefficients for birth weight or family income between the samples for any outcomes measures. Thus, although evidence is weak for these individual models of females, there is little evidence of systematic attrition bias.

Outcome Models with Mother Fixed Effects—In this section we repeat the outcome analysis for models with mother fixed effects. Table 13A presents results for men with brothers. As an intermediate step, the first two columns of the table restrict the sample to the same sample used in the mother fixed effect models, i.e. men who have brothers also remaining in the sample in the indicated year. When restricted to this subsample, the coefficients on birth weight and family income generally become larger in absolute value compared to the individual models in Table 12. Furthermore, some coefficients become larger in the selected 2007 sample and significantly different from the 1991 sample. This hints that attrition may be biasing these coefficients, but our interest is less in this intermediate model and is better directed to the mother fixed effect model.

Birth weight and average family income when the child age is 0 to 16 can vary across the brothers and this allows estimation of their coefficients in a mother fixed effect model. Low birth weight increases the chance of poor health and reduces adult labor income, confirming results in the literature (Smith, 2009; Johnson and Schoeni, 2007). On the right hand side of Table 13A, we observe that the added selection of retaining multiple brothers produces coefficients that are somewhat larger in absolute value in the selected 2007 sample, but the differences are not statistically non-zero. For females with sisters, the mother fixed effect model in Table 13B produces unstable results. That is, coefficients sometimes have counterintuitive signs, or change sign as we go from the full to the selected sample. The only well estimated coefficient indicates that family income increases adult educational achievement, and the impact is somewhat larger in the selected 2007 sample (and significantly different at 10 percent level). But the overall pattern of unstable results and counterintuitive signs for females is not reassuring and weakens firm conclusions.

Robustness Checks—Results for the combined SRC and SEO sample are shown in appendix tables. Results from individual models in Table A3 are qualitatively similar to

[31]Full specifications showing all covariates are in Appendix A1 and A2, for men and women respectively.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

those in the SRC sample and tend to show somewhat larger impacts in the selected sample, and some significant differences in the health models between the selected and complete samples. In Tables A4 and A5, sibling models for the combined SEO+SRC male fixed effect models look similar to the SRC alone. For females, as with the SRC sample, the combined SEO+SRC sample produces some coefficient differences between the complete and retained sample, but results are odd with uneven precision and frequent counterintuitive signs. This suggests a need for further care in using the combined SRC+SEO for fixed effects models with female samples.

As an additional check, the outcome regression models for the SRC sample were also run weighted by 1986 weights. The general pattern of results is stable, but some differences occur in a few models. In the health outcome regression for SRC males and the education regression for education for SRC females, the birth weight coefficients become insignificantly different from zero when weighted. The family income coefficients results change little when weighted. The weighted results tend to have somewhat higher standard errors.

Overall, the conclusion regarding attrition from the outcome regressions is a bit mixed. Substantively, family income and to a less certain extent birth weight appear to affect all three adult outcomes. As for attrition, for the sample of SRC men, where coefficients are measured with greater precision, it appears that coefficients are often larger (in absolute value) in the selected 2007 sample for all the dependent variables. Significant differences would indicate that selective attrition potentially biases the coefficients, but the differences are not statistically different from zero except for the labor income model. This applies to individual models as well as mother fixed effect models. For SRC women, coefficients are less stable across specifications. Coefficients from the selected 2007 sample are sometimes larger and sometimes smaller than the complete sample coefficients, and often imprecisely estimated. Thus it is more difficult to draw a conclusion for females except to say that evidence is weak, but there is not a strong systematic impact of attrition in these models.

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.

9 Retention Probits

A more straightforward test for attrition bias asks whether lagged values of the outcome variables predict future attrition, conditional on the other covariates in the model (FGM). To complement the analysis above, the lag period in this case uses the base year 1991. This section reports probits predicting the probability of responding in 2007 given that the person responded in the 1991 interview. In addition, for sibling models, at least two siblings must survive to the outcome year to estimate a mother fixed effect outcome regression like those in the last section. Consequently we estimate the probability of the event that at least two siblings respond in 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.

Table 14 shows the main results that focus on individual attrition for SRC males and females. To simplify the interpretation, the outcome variable for health is collapsed into a binary indicator that the person experienced poor/fair health at any time during the 1986 to 1991 period. Poor or fair health in that early period is not found to be a significant predictor of attrition in 2007 for males or females, after conditioning on the other covariates. For males, lagged labor income and education are shown to increase retention, conditional on other covariates. For females, although maternal education is shown to increase the probability of retention, none of the lagged outcome variables are significant in predicting

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

later retention. For either gender, the probits are not strong predictors: pseudo R squared is low and few other covariates are significant predictors of attrition[32]. This is consistent with earlier work by FGM. But the significant coefficients on education and earnings for men indicate a potential attrition problem.

Tables 15, 16 and 17 display retention probits relevant to sibling models. Table 15 shows the probability of retaining sibling family groups. That is, consider the subsample of men who have brothers present in the panel in 1991. Of those men, Table 15 shows the probability that the person is retained in 2007 *together with at least one other brother*. Retention of at least two brothers would be necessary for a mother fixed effect model of brothers. The first three columns give results for males with brothers and the right hand columns give results for females with sisters. Among the lagged dependent variables only education in the male sample is a significant predictor, conditional on other covariates.

For sibling models, a separate but closely related question of interest is the retention of sibling pairs. That is, given a sample of unique sibling pairs in 1991, which pairs will be retained in 2007? Table 16 and 17 address this question for the SRC sample. These models are richer than the previous ones in that they include differences between the siblings as well as the level of variables for the oldest sibling. The first three columns show results without sibling differences. The lagged endogenous health variables are not significant for either gender. Lagged labor income for males (not females) and lagged education for females (not males) are significant predictors of retention even after conditioning on the other covariates. Higher levels of mother's education increase retention for sister pairs. The second three columns add variables for the absolute value of the differences between the siblings. For male labor income, siblings who have a higher absolute value of differences in labor income as well as higher labor income are less likely to be retained. Stated differently, siblings who are more similar are more likely to be retained, a result that could weaken models based on sibling differences. Conditional on the variable levels for the older sibling, sibling differences do not predict retention for other dependent variables[33]. For females, the only lagged dependent variable that is statistically non-zero is for education.

Overall, the attrition probit results are a bit mixed. The straightforward individual retention probits suggest that models with lagged education and labor income appear to show some potential attrition bias for males, but the health model does not. The significance of the predictive power of lagged labor income for males carries over into the sibling retention probits for labor income, confirming the analysis from the last section. For females, lagged outcome variables do not consistently predict retention in individual models, again confirming earlier results. In models of sibling retention for females, labor income and health do not predict retention, given other covariates, but education may have an impact.

In short, the retention probits suggest that attrition has less influence on models for females, or at least less conclusive evidence of attrition bias. For men, attrition remains a concern for labor and education outcomes. Lagged health is not a significant predictor or health for either gender, conditional on other covariates.

10 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

[32]Although not relevant for testing attrition bias in the outcome models, it is interesting to note that low birth weight and family income do not predict retention, conditional on the other covariates.

[33]This result is robust to using the signed sibling value of the difference between siblings (older minus younger) instead of the absolute value.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

education, and worse health tend to become non-respondents more frequently. Nonetheless, a comparison of respondents from the PSID and the repeated cross sections of the NHIS shows that the weighted PSID appears to maintain its representativeness along several key dimensions, attesting to the value of the PSID supplied weights. As for health, the PSID consistently has lower responses (worse health) for general health question compared to the NHIS for the child cohort, but we cannot judge the extent to which attrition is responsible because the first health measures in PSID occur midway through the panel after some attrition has occurred. Importantly, the NHIS and the PSID produce similar age-health profiles once one allows for the overall lower level in the PSID. This suggests that attrition is not having a substantial effect on these age-health profiles.

As we turn to intergenerational models with covariates, some caveats apply. Attrition impacts are model specific and this paper investigates a limited number of models, although the paper presents a relevant set of models that resemble those in the intergenerational literature. A limited age range also limits conclusions. We consider a cohort of children aged 0 to 16 in 1968. We observe outcomes in early adulthood when poor health is less likely and there is less variation in health across people, and earnings may not be at permanent levels. Results at older ages might show larger impacts. In a number of models, results vary by gender and are often imprecise. Thus a finding of no statistically significant effects of attrition could reflect low power. Furthermore, because we measure adult outcomes of children at mid-panel after some attrition has occurred, we cannot directly measure the impact of earlier attrition on these outcomes.

Bearing this in mind, we suggest some conclusions. Sibling correlations in outcomes in early adulthood and father-son correlations in earnings are little affected by attrition. The paper finds significant sibling correlations in health, education, and labor measured in 1986 when respondents are in early adulthood. When the sample is restricted to those who are retained in 2007, the selected sample has slightly higher 1986 correlations, but differences between the two samples are not statistically significant. Father-son correlations in earnings behave similarly with no significant difference in correlations between the complete and retained samples.

Tests for attrition bias in the estimation of outcome models generally do not show strong evidence of attrition bias, with some exceptions. In models predicting adult outcomes (health, education, and labor income), 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 slightly 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 (Table 12). Subsamples of females show fewer effects of attrition than males, but also have less stable coefficients across specifications. Retention probits do not show a significant effect of health on retention rates for either gender, although higher education and earnings appear to increase retention for males, after conditioning on other covariates.

Results from sibling models with mother fixed effects are less stable with the selected sample sometimes showing larger and sometimes smaller results, but the differences are marginally significant or insignificant. Results for men show potential attrition effects for labor income but not health or education. Results for women show marginally significant impacts for all outcomes, but the model coefficients often have odd signs and inspire little confidence in those results. Thus some attrition concerns from the individual models appear to carry over to sibling models, but there is not compelling evidence that the added selection of retaining multiple siblings produces sharply higher attrition bias.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Another approach for summarizing the many hypothesis tests would be to reevaluate statistical significance using a Bonferroni correction. Table 18 counts instances where the complete and retained samples produce statistically different estimates of coefficients. It shows the regular results based on a testwise cutoff that considers each test separately and on a Bonferroni adjusted cutoff that controls Type I errors for a group of tests[34]. The grouping of tests for this purpose is a bit arbitrary and should be viewed cautiously. But as a guide to interpretation, the table shows that none of the significant differences between the complete and retained sample stand up to a Bonferroni correction except for the retention probit coefficient on lagged labor income for males. Viewing the testwise results, one gains the impression that labor income and education are more subject to attrition bias than are health models.

The outcomes models and probits offer mixed results but some themes emerge. Results vary by gender and also by the outcome under consideration. Analysts should keep in mind that alternate specifications could produce different impacts. For individual models considered here, the paper finds little evidence of attrition bias for female intergenerational models of parental income and child birth weight impacts on health, education, or earnings outcomes. For sibling models based on sisters, attrition is a potential issue but evidence is weak. For males attention should be given to attrition for adult education and earnings outcomes where we tend to observe stronger intergenerational links in the selected, non-attriting sample. With a possible exception of sibling models for sisters, intergenerational models with covariates that predict adult health outcomes are not likely significantly biased by attrition for either gender for the age ranges and models considered in this paper.

References

Alderman H, Hoddinott J, Kinsey B. Long term consequences of early childhood malnutrition. Oxford Economic Papers. 2006; 58(3):450.

Andreski, P.; McGonagle, K.; Schoeni, R. Panel Study of Income Dynamics Technical Paper Series #07-04. 2007. An Analysis of the Quality of Health Data in the Panel Study of Income Dynamics.

Becketti S, Gould W, Lee L, Welsh F. The Panel Study of Income Dynamics after Fourteen Years: An Evaluation. Journal of Labor Economics. 1988; 6:472–492.

Behrman J. The Intergenerational Correlation Between Adult Earnings and Their Parent's Income: Results from the Michigan Panel of Income Dynamics. Review of Income and Wealth. 36(2):115–127.

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

Case A, Lubotsky D, Paxson C. Economic Status and Health in Childhood: The Origins of the Gradient. American Economic Review. 2002; 92(5):1308–1334.

Conley D, Bennett NG. Is Biology Destiny? Birth Weight and Life Chances. American Sociological Review. 2000; 65(3):458–467.

Conley D, Bennett NG. Birth Weight and Income: Interactions across Generations. Journal of Health & Social Behavior. 2001; 42(4):450–465. [PubMed: 11831142]

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

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

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

Fitzgerald J, Gottschalk P, Moffitt R. An Analysis of Sample Attrition in Panel Data: The Michigan Panel Study of Income Dynamics. The Journal of Human Resources. 1998a; 33(2):251–299.

[34]For multiple hypothesis tests, if an analyst wants to control the probability of a type one error to a specified level α for a group of C tests, the Bonferroni method suggests an approximation that the per test significance level should be α/C.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Fitzgerald J, Gottschalk P, Moffitt R. The Impact of Attrition on the Second Generation of Respondents in the Panel Study of Income Dynamics. The Journal of Human Resources. 1998b; 33(2):300–344.

Gouskova E, Chiteji N, Stafford F. Estimating the Intergenerational Persistence of Lifetime Earnings with Life Course Matching: Evidence from the PSID. Labour Economics. 2010; 17:592–597.

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

Haas S. Health selection and the process of social stratification: The effect of childhood health on socioeconomic attainment. Journal of Health and Social Behavior. 2006; 47(4):339–354. [PubMed: 17240924]

Haider S, Solon G. Life-Cycle Variation in the Association between Current and Lifetime Earnings. American Economic Review. 2006; 96(4):1308–1320.

Halliday, TJ.; Kimmitt, M. IZA Discussion Paper. 2008. Selective Migration and Health.

Heckman JJ. Sample Selection Bias as a Specification Error. Econometrica. 1979; 47(1):153–161.

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

Minnesota Population Center Integrated Health Interview Series. University of Minnesota; http://www.ihis.us/ihis/

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

Lillard LA, Panis CWA. Panel Attrition from the Panel Study of Income Dynamics: Household Income, Marital Status, and Mortality. Journal of Human Resources. 1998; 33(2):437–457.

Manski CF. Anatomy of the Selection Problem. The Journal of Human Resources. 1989; 24(3):343.

Meer J, Miller DL, Rosen HS. Exploring the Health-Wealth Nexus. Journal of Health Economics. 2003; 22(5):713–730. [PubMed: 12946455]

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

Smith JP. The Impact of Socioeconomic Status on Health over the Life-Course. Journal of Human Resources. 2007; 42(4):739–764.

Smith JP. The Impact of Childhood Health on Adult Labor Market Outcomes. Review of Economics and Statistics. 2009; 91(3):478–489.

Solon G, Corcoran M, Gordon R, Laren D. Sibling and Intergenerational Correlations in Welfare Program Participation. NBER Working Paper No. 2334. 1987

Solon G, Corcoran M, Gordon R, Laren D. A Longitudinal Analysis of Sibling Correlations in Economic Status. Journal of Human Resources. 1991; 26(3):509–534.

Solon. Intergenerational Income Mobility in the United States. American Economic Review. 1992; 82:393–408.

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

Wooldridge, JM. Econometric Analysis of Cross Section and Panel Data. Cambridge MA: MIT Press; 2002.

Ziliak, James P.; Kniesner, Thomas J. The Importance of Sample Attrition in Life Cycle Labor Supply Estimation. Journal of Human Resources. 1998; 33(2):507–530.

Data Appendix

	Sample and data definitions
Table 1,2,3	Sample members aged 0 to 16 in 1968.
Table 4,5	Sample members aged 0 to 16 in 1968 who are present in 1986.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

	Sample and data definitions
Table 6	Head or Wife of family with child aged 0 to 16 in 1968.
Table 7, 8, 9, 10	Child[1] aged 0 to 16 in 1968 present in indicated years and household head or wife in indicated years.
Table 11	Child aged 0 to 16 in 1968, SRC only, present in 1991. Labor Income averaged over son's age 25 to 34
Table 12,14, A1, A2	Child aged 0 to 16 in 1968, SRC only, present in 1991. Health measured in each year 1986 to 1991. Education measured in for respondents at least age 24 in 1991. Labor Income averaged over respondent's age 25 to 34.
Table 13A, B	Child aged 0 to 16 in 1968, SRC only, present in 1991. For sibling sample, mother must be identified, and at least one sibling of correct gender present. Health measured in each year 1986 to 1991. Education measured in for respondents at least age 24 in 1991. Labor Income averaged over respondent's age 25 to 34.
Table 15	Child aged 0 to 16 in 1968, SRC only, present in 1991. For sibling sample, mother must be identified, and at least one sibling of correct gender present. Health is single measure of whether ever had poor or fair health during 1986 to 1991. Education measured in for respondents at least age 24 in 1991. Labor Income averaged over respondent's age 25 to 34.
Table 16,17	Child aged 0 to 16 in 1968, SRC only, present in 1991. For sibling pairs, mother must be identified, and at least one sibling of correct gender present. Each pair is one observation. Both siblings must be aged 0 to 16 in 1968. Health is single measure of whether ever had poor or fair health during 1986 to 1991. Education measured in for respondents at least age 24 in 1991. Labor Income averaged over respondent's age 25 to 34.
Table A3	Child aged 0 to 16 in 1968, SRC and SEO sample, present in 1991. Health measured in each year 1986 to 1991. Education measured in for respondents at least age 24 in 1991. Labor Income averaged over respondent's age 25 to 34.
Table A4, A5	Child aged 0 to 16 in 1968, SRC and SEO sample, present in 1991. For sibling sample, mother must be identified, and at least one sibling of correct gender present. Health measured in each year 1986 to 1991. Education measured in for respondents at least age 24 in 1991. Labor Income averaged over respondent's age 25 to 34.

[1]Child in 1968 means that person is listed as child of head who was original PSID member in 1968.

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

B E J Econom Anal Policy. Author manuscript; available in PMC 2012 February 23.

Appendix Table A1

Regression of Adult Outcomes on Child Background Characteristics, SRC Sample Males

Dependent Variable	Bad Health	Bad Health		Education	Education		Labor Income	Labor Income	
Sample	Complete 1991	Retained in 2007	Diff p value	Complete 1991	Retained in 2007	Diff p value	Complete 1991	Retained in 2007	Diff p value
Low birth weight (<5.5 lbs)	.025** (0.012)	.032** (0.016)	.28	−0.124 (0.321)	−.0538 (0.388)	.73	−0.174* (0.090)	−0.328*** (0.100)	.01
Log Family income, child 0 to 16	−.0018 (0.0037)	−0.0059 (0.0039)	.16	1.28*** (0.160)	1.372*** (0.189)	.42	0.251*** (0.075)	0.285*** (0.082)	.53
Mother education =12	−0.011** (0.0053)	−0.0016 (0.0049)		.729*** (0.175)	.734*** (0.197)		.218** (.0923)	.121 (.0937)	
=13/15	−0.015*** (0.0048)	−0.0060 (0.0040)		1.61*** (0.253)	1.504*** (0.278)		.219** (.120)	.154 (.118)	
>=16	−0.014** (0.0055)	−0.0026 (0.0047)		1.74*** (0.281)	1.48*** (0.323)		.299** (.139)	.226 (.152)	
Age	−0.0088* (0.0047)	−0.0078 (0.0051)		0.042 (0.246)	0.22 (0.276)		0.363 (0.251)	0.488* (0.281)	
Birth order number	0.0018 (0.0013)	0.0026 (0.0016)		−0.294*** (0.057)	−0.307*** (0.0645)		−.076*** (0.0197)	−0.083*** (0.0211)	
Mom married in 1968	−0.0047 (0.00074)	−0.010 (0.0087)		−0.288 (0.285)	−0.346 (0.0323)		−.0538 (.143)	−.0262 (.162)	
Age squared	0.017** (0.00008)	0.00014 (0.000088)		−0.000503 (0.0039)	−0.000344 (0.00433)		−0.0063 (0.0042)	−0.00831* (0.00477)	
Black	0.0048 (0.010)	0.014 (0.013)		0.407 (0.279)	0.531 (0.334)		−.381*** (.126)	−.235 (.135)	
Hisp1968	−0.0025 (0.0028)	0.0045 (0.0060)		.259 (0.6889)	1.059* (0.590)		.0666 (.136)	.123 (.250)	
Observations	3799	2906		728	564		449	354	

Notes: Dependent Variables: Bad health is fair or poor health on 5 point scale, measured each year 1986 to 1991. Education is years of education in 1991 when respondent is at least age 24. Labor Income is averaged over ages 25 – 34. Robust standard errors in parentheses.

* significant at 10%;

** significant at 5%;

*** significant at 1%. PSID SRC sample: Child of head aged 0 to 16 in 1968 with known mother. 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.

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

Appendix Table A2

Regression of Adult Outcomes on Child Background Characteristics, SRC Sample Females

Dependent Variable	Bad Health	Bad Health		Education	Education		Labor Income	Labor Income	
Sample	Complete 1991	Retained in 2007	Diff p value	Complete 1991	Retained in 2007	Diff p value	Complete 1991	Retained in 2007	Diff p value
Low birth weight (<5.5 lbs)	−0.00096 (0.0043)	−0.0041 (0.0043)	.29	−0.355 (0.290)	−0.455 (0.339)	.49	0.364** (0.172)	0.300 (0.191)	.46
Log Family income, child 0 to 16	.00098 (0.0024)	.0018 (0.0029)	.31	1.09*** (0.145)	1.09*** (0.159)	.94	0.482*** (0.122)	0.475*** (0.134)	.91
Mother education =12	−0.0351 (0.00290)	−0.00354 (0.00364)		.9349*** (0.1586)	.9633*** (0.1795)		.350** (.139)	.222 (.143)	
=13/15	−0.000363 (0.00528)	−0.00122 (0.00614)		1.897*** (0.2504)	1.839*** (0.2682)		.389** (.181)	.330* (.182)	
>=16	−0.00721 (0.00442)	−0.0123*** (0.00399)		2.312*** (0.2560)	2.374*** (0.2858)		.715*** (.190)	.569*** (.211)	
Age	0.00263 (0.00318)	0.00183 (0.00397)		−0.616*** (0.2211)	−0.5545** (0.2425)		0.0022 (0.420)	0.346 (0.444)	
Birth order number	−0.00022 (0.000919)	−0.000286 (0.0011)		−0.0562 (0.05122)	−0.0679*** (0.0588)		−.0300 (0.039)	−0.019*** (0.0393)	
Mom married in 1968	−0.00342 (0.00538)	−0.00726 (0.00675)		−0.5035* (0.2883)	−0.4586 (0.03302)		−0.418 (.285)	−.340 (.302)	
Age squared	−0.0000418 (0.0000522)	−0.0000281 (0.0000647)		0.010379*** (0.00349)	0.00943*** (0.003829)		−0.00082 (0.0072)	−0.0067 (0.0076)	
Black	−0.0033 (0.00384)	−0.00705* (0.00427)		0.02867 (0.2288)	0.3328 (0.2723)		0.186 (.193)	0.126 (.210)	
Hisp1968	−0.0071*** (0.00271)	−0.00756 (0.00324)		−0.73851 (0.7831)	−0.7763* (0.7811)		−.069 (.241)	−0.153 (.279)	
Observations	4137	3249		719	574		469	379	

Notes: Dependent Variables: Bad health is fair or poor health on 5 point scale, measured each year 1986 to 1991. Education is years of education in 1991 when respondent is at least age 24. Labor Income is averaged over ages 25 – 34. Robust standard errors in parentheses.

* significant at 10%;

** significant at 5%;

*** significant at 1%. PSID SRC sample: Child of head aged 0 to 16 in 1968 with known mother. 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.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Appendix Table A3

Outcome Regressions: SRC + SEO Sample Individuals

	Males			Females		
A. Bad Health sample (dependent variable)						
	Complete 1991	**Retained in 2007**	**P value for Difference**	**Complete 1991**	**Retained in 2007**	**P value for Difference**
Low birth weight	0.014* (0.0075)	0.040*** (0.015)	0.004	−0.00017 (0.0037)	−0.0039 (0.0034)	0.29
Log Family income, child 0 to 16	.0030 (0.0033)	.001 (0.005)	0.56	0.0047* (0.0025)	0.00011 (0.0026)	0.03
Observations	7337	3891		8862	5200	
B. Individual Education (dependent variable)						
Low birth weight	−0.146 (0.190)	−0.066 (0.302)	0.70	−0.329** (0.170)	−0.340 (0.238)	0.94
Log Family income, child 0 to 16	1.02*** (0.101)	1.10*** (0.151)	0.50	0.720*** (0.010)	0.769*** (0.125)	0.55
Observations	1451	772		1561	934	
C. Labor (dependent variable)						
Low birth weight	−0.219** (0.095)	−0.261*** (0.085)	0.65	0.023 (0.126)	0.177 (0.155)	0.11
Log Family income, child 0 to 16	0.382*** (0.063)	0.331*** (0.085)	0.47	0.503*** (0.094)	0.555*** (0.121)	0.50
Observations	950	493		1026	637	

Notes: Dependent Variables: Bad health is fair or poor health on 5 point scale, measured each year 1986 to 1991. Education is years of education in 1991 when respondent is at least age 24. Labor Income is averaged over ages 25 – 34. Robust standard errors in parentheses.

*significant at 10%;

**significant at 5%;

***significant at 1%. PSID SRC sample: Child of head aged 0 to 16 in 1968 with known mother. 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.

Appendix Table A4

Outcome Regressions: SRC + SEO Sample Males with Brothers

	Individuals with Brothers			Mother Fixed Effects		
A. Bad Health sample (dependent variable)						
	Complete 1991	**Retained in 2007**	**P value for Difference**	**Complete 1991**	**Retained in 2007**	**P value for Difference**
Low birth weight	0.031** (0.014)	0.093*** (0.031)	0.0004	0.056 (0.015)	0.097*** (0.028)	0.009
Log Family income, child 0 to 16	0.007 (0.005)	0.018** (0.008)	0.08	0.031** (0.018)	0.067** (0.034)	0.17
Observations	4076	1894		4076	1894	
B. Individual Education (dependent variable)						

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

A. Bad Health sample (dependent variable)						
	Individuals with Brothers			Mother Fixed Effects		
	Complete 1991	Retained in 2007	P value for Difference	Complete 1991	Retained in 2007	P value for Difference
Low birth weight	0.072 (0.297)	0.483 (0.498)	0.19	−0.010 (0.244)	−0.207 (0.383)	0.95
Log Family income, child 0 to 16	1.07*** (0.157)	1.29*** (0.250)	0.31	0.584 (0.464)	0.786 (0.797)	0.75
Observations	763	345		763	345	
C. Labor (dependent variable)						
Low birth weight	−0.241** (0.112)	−0.401*** (0.096)	0.15	−0.208 (0.135)	−0.363*** (0.064)	0.27
Log Family income, child 0 to 16	0.527*** (0.081)	0.436*** (0.095)	0.36	0.431 (0.356)	0.0246 (0.517)	0.47
Observations	518	229		518	229	

Notes: Dependent Variables: Bad health is fair or poor health on 5 point scale, measured each year 1986 to 1991. Education is years of education in 1991 when respondent is at least age 24. Labor Income is averaged over ages 25 – 34. Robust standard errors in parentheses.

*significant at 10%;

**significant at 5%;

***significant at 1%. PSID SRC sample: Child of head aged 0 to 16 in 1968 with known mother. 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.

Appendix Table A5

Outcome Regressions: SRC + SEO Sample Females with Sisters

A. Bad Health sample (dependent variable)						
	Individuals with Sisters			Mother Fixed Effects		
	Complete 1991	Retained in 2007	P value for Difference	Complete 1991	Retained in 2007	P value for Difference
Low birth weight	−0.0035 (0.0041)	−0.0069*** (0.0018)	0.40	−0.019*** (0.068)	−0.0014** (0.00062)	0.0092
Log Family income, child 0 to 16	0.0064 (0.0037)	−.0043 (0.0037)	0.002	0.079*** (0.021)	0.011 (0.010)	0.0011
Observations	5535	2745		5535	2745	
B. Individual Education (dependent variable)						
Low birth weight	−0.084 (0.188)	−0.215 (0.267)	0.54	0.610*** (0.185)	0.164 (0.237)	0.02
Log Family income, child 0 to 16	0.795*** (0.122)	0.828*** (0.166)	0.78	0.265 (0.324)	1.16*** (0.408)	0.0061
Observations	943	467		943	467	
C. Labor (dependent variable)						
Low birth weight	0.077 (0.145)	0.331** (0.131)	0.08	0.174 (0.168)	0.392** (0.158)	0.18

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

A. Bad Health sample (dependent variable)

	Individuals with Sisters			Mother Fixed Effects		
	Complete 1991	**Retained in 2007**	**P value for Difference**	**Complete 1991**	**Retained in 2007**	**P value for Difference**
Log Family income, child 0 to 16	0.513*** (0.127)	0.642*** (0.166)	0.31	0.350 (0.828)	1.59 (1.13)	0.07
Observations	654	340		654	340	

Notes: Dependent Variables: Bad health is fair or poor health on 5 point scale, measured each year 1986 to 1991. Education is years of education in 1991 when respondent is at least age 24. Labor Income is averaged over ages 25 – 34. Robust standard errors in parentheses.

* significant at 10%;

** significant at 5%;

*** significant at 1%. PSID SRC sample: Child of head aged 0 to 16 in 1968 with known mother. 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.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

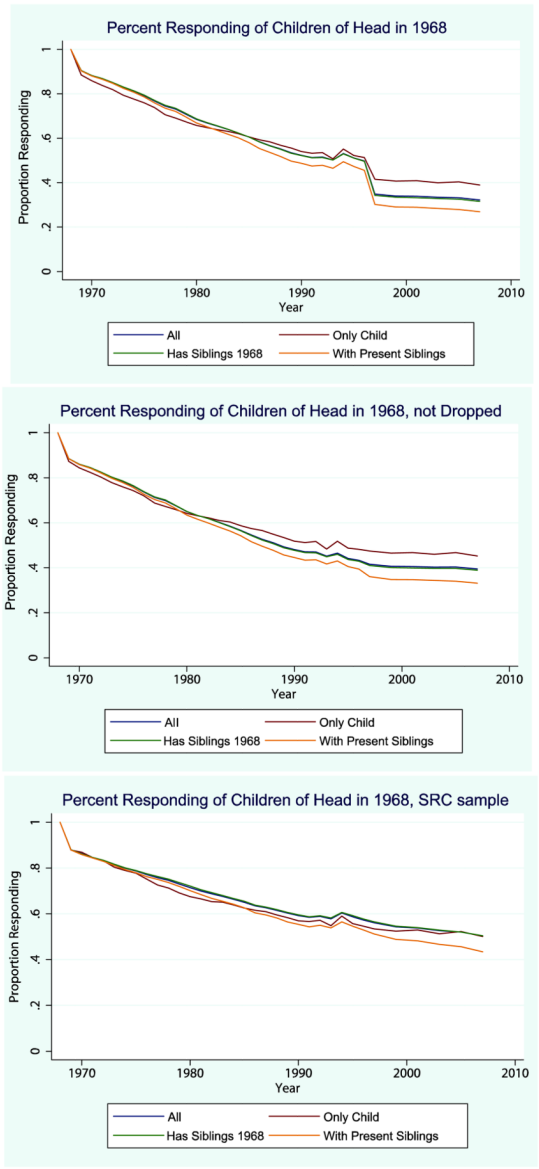

Figure 1.
Percent Responding in PSID: Children Aged 0 to 16 in 1968
Source: Author's computation from PSID.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

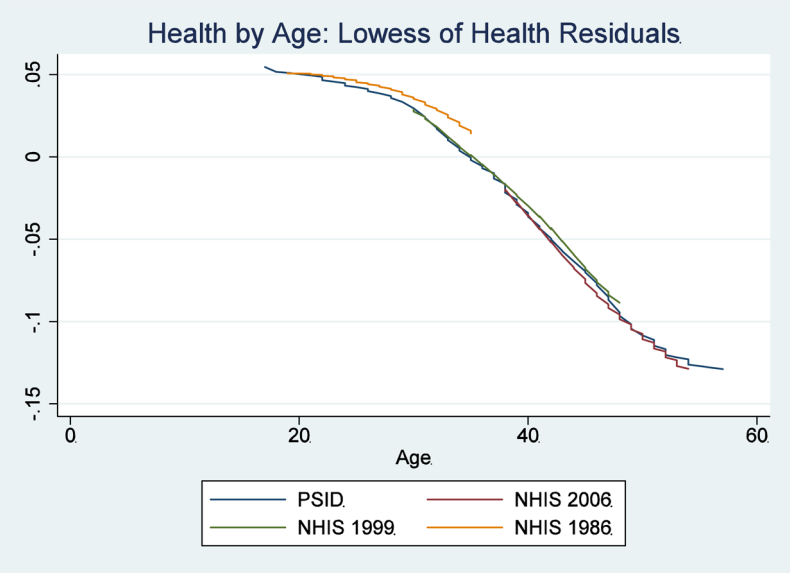

Figure 2.
Health by Age in PSID and NHIS
Source: Author's computation. Child cohort as it ages from 0–16 in 1968, to age 18 to 34 in 1986, to age 31 to 47 in 1999, and to age 39–55 in 2006/07. Lowess of residuals removing means by race and Hispanic origin interacted with gender, adjusted to common point at age 19 for both surveys. PSID cohort aged 0 to 16 in 1968. NHIS surveys 1986, 1999, 2006.

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

B E J Econom Anal Policy. Author manuscript; available in PMC 2012 February 23.

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 sibling in 1968**	**Presently has Siblings**
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 sibling in 1968 %	Presently has Siblings
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 sibling in 1968	Presently has Siblings
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.

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

B E J Econom Anal Policy. Author manuscript; available in PMC 2012 February 23.

Table 2

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

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
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
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

Notes: Unweighted counts.

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

B E J Econom Anal Policy. Author manuscript; available in PMC 2012 February 23.

Table 3

1968 Characteristics by Attrition Status in 2007

Sample: Children 0–16 in 1968					
Variable	**Complete 1968 Sample**	**Retained 2007**	**Out in 2007 (NR, not dead or SEO cut)**	**Out in 2007 by death**	**Out in 2007 by SEO cut**
Characteristics 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 to 11	39.44	24.97	45.66	46.61	54.62

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

Sample: Children 0–16 in 1968					
Variable	**Complete 1968 Sample**	**Retained 2007**	**Out in 2007 (NR, not dead or SEO cut)**	**Out in 2007 by death**	**Out in 2007 by SEO cut**
12	44.17	50.75	42.33	40.9	32.45
13 to 15	9.83	14.9	7.23	6.82	8.52
16 or more	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
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 to 11	40.64	20.41	47.03	49.35	46.57
12	31.11	36.61	28.81	29.53	26.53
13 to 15	13.55	16.78	13.22	10.71	20.55
16 or more	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, aged 0–16 in 1968. Weighted by 1968 person weight. SEO cut is 1997 sample reduction of SEO subsample. NR means non-response.

B E J Econom Anal Policy. Author manuscript; available in PMC 2012 February 23.

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

Table 4

1986 Characteristics of Children as Adults by Attrition Status in 2007

	Complete 1986 Sample	Retained in 2007	Out in 2007 (NR, not dead or SEO cut)	Out in 2007 by death	Out in 2007 by 97 SEO cut
Characteristics 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 to 11	16.4	13.04	20.29	22.74	23.34
12	40.49	38.69	41.55	52.44	44.12
13 to 15	24.96	26.85	23.31	19.06	20.68
16 or more	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

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

	Complete 1986 Sample	Retained in 2007	Out in 2007 (NR, not dead or SEO cut)	Out in 2007 by death	Out in 2007 by 97 SEO cut
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 to 11	28.8	18.94	33.13	38.59	45.54
12	45.09	49.1	44.82	40.84	37.04
13 to 15	14.95	17.87	13.72	12.56	8.46
16 or more	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
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

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

	Complete 1986 Sample	Retained in 2007	Out in 2007 (NR, not dead or SEO cut)	Out in 2007 by death	Out in 2007 by 97 SEO cut
Divorced/Separated	6.38	7.51	4.84	4.27	12.89
Never Married	0.02	0	0	0	0.28
Education					
0 to 11	33.34	22.45	36.11	42.47	45.61
12	30.72	29.13	25.73	34	31.31
13 to 15	15.03	17.37	17.76	12.19	11.42
16 or more	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, aged 0–16 in 1968, age 18–34 in 1986. Weighted by 1968 person weight. SEO cut is 1997 sample reduction of SEO subsample. NR means non-responding.

B E J Econom Anal Policy. Author manuscript; available in PMC 2012 February 23.

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

B E J Econom Anal Policy. Author manuscript; available in PMC 2012 February 23.

Table 5A

1986 Characteristics by Attrition Status for Sibling Pairs

		Status in 2007			
Variable	**Complete in 1986**	**Both retained 2007**	**Either out, Nonresponse**	**Either out by death**	**Either out by SEO cut**
Characteristics of Child (Older sibling)					
Sample	5699	2000	1635	373	1691
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 (Years Completed)					
0 to 11	17.18	12.05	19.91	27.54	21.56
12	39.05	38.5	36.07	42.77	42.63
13 to 15	23.52	24.43	25.91	19.14	20.05
16 or more	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
	N for health				
Sibling Correlations in 1986	4511	1692	1178	292	1349
Good health	0.19	0.23	0.19	−0.07	0.12

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

		Status in 2007			
Variable	**Complete in 1986**	**Both retained 2007**	**Either out, Nonresponse**	**Either out by death**	**Either out by SEO cut**
test complete=both retained					
p-value		0.02			
Labor Income	0.22	0.23	0.25	0.1	0.14
test complete=both retained					
p-value		0.83			
Education	0.43	0.44	0.45	0.18	0.35
test complete=both retained					
p-value		0.63			

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 log of average over age 25–34; Education is years of education at age 24.

B E J Econom Anal Policy. Author manuscript; available in PMC 2012 February 23.

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

B E J Econom Anal Policy. Author manuscript; available in PMC 2012 February 23.

Table 5B

1986 Characteristics by Attrition Status for Male Sibling Pairs

		Status in 2007			
Variable	**Complete in 1986**	**Both Retained 2007**	**Either out, Nonresponse**	**Either out by death**	**Either out by SEO cut**
Characteristics of Child (Older sibling)					
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 (Years Completed)					
0 to 11	18.1	12.45	16.32	30.43	27.02
12	39.11	38.94	34.02	44.92	42.72
13 to 15	22.42	23.08	28.34	17.58	16.52
16 or more	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***

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

		Status in 2007			
Variable	**Complete in 1986**	**Both Retained 2007**	**Either out, Nonresponse**	**Either out by death**	**Either out by SEO cut**
test complete=both retained					
p-value		0.51			
Labor Income	.38***(.016)	.41***	.32***	.18**	.3***
test complete=both retained					
p-value		0.27			
Education	.43***(.025)	.48***	.45***	−0.01	.26***
test complete=both retained					
p-value		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 log of 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)

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

B E J Econom Anal Policy. Author manuscript; available in PMC 2012 February 23.

Table 5C

1986 Characteristics by Attrition Status for Female Sibling Pairs

		Status in 2007			
Variable	**Complete in 1986**	**Both Retained 2007**	**Either out, Nonresponse**	**Either out by death**	**Either out by SEO cut**
Characteristics of Child (Older Sibling)					
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 (Years Completed)					
0 to 11	16.61	13.76	19.99	24.38	16.81
12	40.25	37.56	39.6	40.96	46.49
13 to 15	23.8	25.26	24.99	15.67	20.63
16 or more	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	537	360	62	371
Good health	.32*** (.020)	.36***	.26***	.21*	.30***

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

		Status in 2007			
Variable	**Complete in 1986**	**Both Retained 2007**	**Either out, Nonresponse**	**Either out by death**	**Either out by SEO cut**
test complete=both retained					
p-value		0.15			
Labor Income	.26*** (.021)	.23***	.26***	.32***	.31***
test complete=both retained					
p-value		0.31			
Education	.43*** (.021)	.46***	.40***	.21*	.43***
test complete=both retained					
p-value		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 log of 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)

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Table 6

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.

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

B E J Econom Anal Policy. Author manuscript; available in PMC 2012 February 23.

Table 7

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	11795	2739	21430	2605	15238
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						

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

Study Year	PSID 1986	NHIS 1986	PSID 1999	NHIS 1999	PSID 2007	NHIS 2006
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 aged 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).

B E J Econom Anal Policy. Author manuscript; available in PMC 2012 February 23.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Table 8

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	30.96	27.66	19.96
Very Good	37.39	36.07	37.20
Good	24.84	27.40	29.54
Fair	6.06	7.27	9.96
Poor	0.75	1.60	3.35

PSID-SRC only		**Survey year**	
Health status	1986	1999	2007
Age	18–34	31–47	38–54
Excellent	32.07	28.44	20.61
Very Good	38.57	36.90	37.75
Good	23.08	26.48	29.09
Fair	5.67	6.67	9.30
Poor	0.61	1.51	3.25

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

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Table 9

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	8963	6311
White female	0.74	0.72	0.63
	6521	9441	6707
Black male	0.64	0.64	0.52
	1069	1375	1141
Black female	0.56	0.59	0.47
	1499	1885	1562
Other male	0.72	0.65	0.64
	379	1032	698
Other female	0.64	0.60	0.61
	418	1155	730

PSID SEO+SRC		Survey year	
Race-sex	1986	1999	2007
Age	18–34	31–47	38–54
White male	0.76	0.69	0.63
	830	733	697
White female	0.67	0.67	0.57
	969	748	714
Black male	0.58	0.46	0.52
	537	368	376
Black female	0.48	0.40	0.37
	779	621	640
Other male	0.46	0.70	0.51
	11	24	10
Other female	0.56	0.46	0.43
	32	48	29

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

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Table 10

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

Dependent Variable	Model 1 good health	Model 2 good health	Model 1 Weighted good health	Model 2 Weighted good health
psid	−0.106*** (0.009)	−0.062*** (0.009)	−0.056*** (0.011)	−0.050*** (0.011)
Year 1999	−0.041*** (0.005)	0.066*** (0.007)	−0.037*** (0.005)	0.064*** (0.008)
Year 2007	−0.136*** (0.005)	0.024*** (0.009)	−0.126*** (0.006)	0.028** (0.011)
Psid × 1999	−0.0086 (0.014)	−0.035*** (0.013)	−0.011 (0.016)	−0.026* (0.016)
Psid × 2007	0.027*** (0.014)	0.010 (0.014)	0.011 (0.017)	0.007 (0.017)
Age		−0.008*** (0.0004)		−0.008*** (0.0005)
White female		−0.033*** (0.004)		−0.032*** (0.005)
Black male		−0.136*** (0.008)		−0.149*** (0.010)
Black female		−0.217*** (0.007)		−0.216*** (0.008)
Other male		−0.047*** (0.011)		−0.047*** (0.014)
Other female		−0.109*** (0.011)		−0.105*** (0.014)
Hispanic male		−0.108*** (0.008)		−0.094*** (0.010)
Hispanic female		−0.115*** (0.008)		−0.113*** (0.010)
Constant	0.733*** (0.003)	0.989*** (0.011)	0.742*** (0.004)	0.997*** (0.014)
Observations	65417	65417	65216	65216
R-squared	0.02	0.05	0.02	0.04

Notes: Dependent variable is Good or Excellent Health on a 5 point scale. 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%

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Table 11

Father Son Elasticities in Labor Income

	Log Average Son's Labor Income age 25 to 34		
Father's Labor Income	**Respondent 1991**	**Respondent 2007**	**p-value for difference in samples**
Log Family income, when son age 0 to 16	.46 (.07) [382]	.45 (.07) [306]	.86
Log average labor income, age 25 to 34	.51 (.09) [165]	.51 (.10) [134]	.97
Log average labor income, age 35 to 44	.34 (.06) [335]	.37 (.06) [272]	.42
	Average of Log Son's Labor Income age 25 to 34		
Average of log labor income, age 25 to 34	.47 (.07) [165]	.47 (.08) [134]	.90
Average of log labor income, age 35 to 44	.34 (.06) [335]	.36 (.07) [272]	.53

Notes: Robust standard errors in parentheses. Sample size in brackets. Model includes quadratics in son's and father's ages. PSID SRC sample of sons aged 0–16 in 1968.

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

B E J Econom Anal Policy. Author manuscript; available in PMC 2012 February 23.

Table 12

Regression of Adult Outcomes on Child Background Characteristics

Dependent Variable	Bad Health	Bad Health		Education	Education		Labor Income	Labor Income	
Sample	**Complete 1991**	**Retained in 2007**	**Diff p value**	**Complete 1991**	**Retained in 2007**	**Diff p value**	**Complete 1991**	**Retained in 2007**	**Diff p value**
A. SRC Sample Males									
Low birth weight (<5.5 lbs)	.025** (0.012)	.032** (0.016)	.28	−0.124 (0.321)	−.0538 (0.388)	.73	−0.174* (0.090)	−0.328*** (0.100)	.01
Log Family income, child 0 to 16	−.0018 (0.0037)	−0.0059 (0.0039)	.16	1.28*** (0.160)	1.372*** (0.189)	.42	0.251*** (0.075)	0.285*** (0.082)	.53
Mother education =12	−0.011** (0.0053)	−0.0016 (0.0049)		.729*** (0.175)	.734*** (0.197)		.218** (.0923)	.121 (.0937)	
=13/15	−0.015*** (0.0048)	−0.0060 (0.0040)		1.61*** (0.253)	1.504*** (0.278)		.219** (.120)	.154 (.118)	
>=16	−0.014** (0.0055)	−0.0026 (0.0047)		1.74*** (0.281)	1.48*** (0.323)		.299** (.139)	.226 (.152)	
Observations	3799	2906		728	564		449	354	
R-squared	0.01	0.02		.25	.24		.16	.15	
B. SRC Sample Females									
Low birth weight (<5.5 lbs)	−0.00096 (0.0043)	−0.0041 (0.0043)	.29	−0.355 (0.290)	−0.455 (0.339)	.49	0.364** (0.172)	0.300 (0.191)	.46
Log Family income, child 0 to 16	.00098 (0.0024)	.0018 (0.0029)	.31	1.09*** (0.145)	1.09*** (0.159)	.94	0.482*** (0.122)	0.475*** (0.134)	.91
Mother education =12	−0.0351 (0.00290)	−0.00354 (0.00364)		.9349*** (0.1586)	.9633*** (0.1795)		.350** (.139)	.222 (.143)	
=13/15	−0.000363 (0.00528)	−0.00122 (0.00614)		1.897*** (0.2504)	1.839*** (0.2682)		.389** (.181)	.330* (.182)	
>=16	−0.00721 (0.00442)	−0.0123*** (0.00399)		2.312*** (0.2560)	2.374*** (0.2858)		.715*** (.190)	.569*** (.211)	
Observations	4137	3249		719	574		469	379	
R-squared	0.01	0.02		.30	.29		.11	.10	

Notes: Dependent Variables: Bad health is fair or poor health on 5 point scale, measured each year 1986 to 1991. Education is years of education in 1991 when respondent is at least age 24. Labor Income is averaged over ages 25 – 34. Robust standard errors in parentheses.

* significant at 10%;

** significant at 5%;

*** significant at 1%. PSID SRC sample: Child of head aged 0 to 16 in 1968 with known mother. 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.

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

Table 13A

Outcome Regressions, SRC Males with Brothers

	Individuals with Brothers			With Mother Fixed Effects		
A. Fair/Poor health (dependent variable)						
	Complete 1991	**Retained in 2007**	**P value for Difference**	**Complete 1991**	**Retained in 2007**	**P value for Difference**
Low birth weight	0.054** (0.025)	0.076** (0.034)	0.03	0.049* (0.024)	0.066* (0.036)	0.19
Log Family income 0 to 16	0.0074 (0.0056)	−0.00021 (0.0043)	0.20	0.076** (0.030)	0.104** (0.047)	0.17
Observations	1988	1420		1988	1420	
B. Individual Education (dependent variable)						
Low birth weight	0.848* (0.445)	1.01* (0.503)	0.54	0.412 (0.290)	0.221 (0.429)	0.36
Log Family income 0 to 16	1.48*** (0.263)	1.96*** (0.304)	0.063	0.228 (0.804)	0.639 (0.105)	0.56
Observations	365	257		365	257	
C. Labor Income (dependent variable)						
Low birth weight	−0.361** (0.141)	−0.472*** (0.129)	0.35	−0.373*** (0.138)	−0.465*** (0.105)	0.47
Log Family income 0 to 16	0.455*** (0.087)	0.553*** (0.105)	0.20	0.216 (0.500)	.874 (0.577)	0.23
Observations	237	171		237	171	

Notes: Dependent Variables: Bad health is fair or poor health on 5 point scale, measured each year 1986 to 1991. Education is years of education in 1991 when respondent is at least age 24. Labor Income is averaged over ages 25 – 34. Robust standard errors in parentheses.

* significant at 10%;

** significant at 5%;

*** significant at 1%. PSID SRC sample: Child of head aged 0 to 16 in 1968 with known mother. 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.

B E J Econom Anal Policy. Author manuscript; available in PMC 2012 February 23.

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

Table 13B

Outcome Regressions, SRC Females with Sisters

	Individuals with Sisters			With Mother Fixed Effects		
	Completed 1991	**Retained in 2007**	**P value for Difference**	**Complete 1991**	**Retained in 2007**	**P value for Difference**
A. Fair/Poor health (dependent variable)						
Low birth weight	−0.0015 (0.0045)	−0.0045** (0.0019)	0.48	−0.0018* (0.00096)	−0.0019* (0.0010)	0.89
Log Family income 0 to 16	−.0022 (0.0026)	−.0014 (0.0031)	0.63	.0020 (0.0083)	0.014* (0.0081)	0.09
Observations	2312	1584		2312	1584	
B. Individual Education (dependent variable)						
Low birth weight	.133 (0.333)	−0.182 (0.370)	0.22	0.308 (0.342)	−.021 (0.317)	0.16
Log Family income 0 to 16	1.24*** (0.196)	1.16*** (0.221)	0.54	1.68*** (0.487)	2.13*** (0.610)	0.09
Observations	393	269		393	269	
C. Labor Income (dependent variable)						
Low birth weight	0.471** (0.188)	0.486*** (0.171)	0.93	0.516*** (0.178)	0.309 (0.205)	0.10
Log Family income 0 to 16	0.538*** (0.163)	0.659*** (0.199)	0.36	−0.729 (0.843)	0.143 (.877)	0.13
Observations	276	191		276	191	

Notes: Dependent Variables: Bad health is fair or poor health on 5 point scale, measured each year 1986 to 1991. Education is years of education in 1991 when respondent is at least age 24. Labor Income is averaged over ages 25 – 34. Robust standard errors in parentheses.

* significant at 10%;

** significant at 5%;

*** significant at 1%. PSID SRC sample: Child of head aged 0 to 16 in 1968 with known mother. 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.

B E J Econom Anal Policy. Author manuscript; available in PMC 2012 February 23.

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

Table 14

Probits for Being Sample Respondent in 2007, SRC

	SRC Males			SRC Females		
Has bad health in 86–91	−0.155 (0.117)			−0.0421 (0.0939)		
Individual Education		0.0160** (0.00821)			0.0121 (0.00839)	
Log Labor Income Averaged, age 24–35			0.0650* (0.0291)			0.00966 (0.0162)
Low birth weight(d)	−0.108 (0.0808)	−0.0874 (0.127)	−0.120 (0.0964)	0.00884 (0.0507)	−0.00731 (0.0541)	0.0245 (0.0590)
Log Family income Averaged, age 0–16	0.0416 (0.0409)	0.0279 (0.0711)	0.0651 (0.0511)	0.0231 (0.0307)	0.0170 (0.0322)	0.0299 (0.0438)
Age	0.0293 (0.0474)	−0.0337 (0.0969)	0.152 (0.172)	0.0171 (0.0438)	0.0267 (0.0495)	0.116 (0.150)
Birth Order Number	0.00454 (0.0121)	0.0103 (0.0191)	0.0216 (0.0139)	−0.00257 (0.0119)	−0.000555 (0.0123)	−0.00455 (0.0129)
Mother's Education (<12 omitted)						
=12 years(d)	−0.0321 (0.0381)	−0.0445 (0.0384)	−0.0545 (0.0488)	0.0169 (0.0357)	−0.00135 (0.0374)	−0.0496 (0.0471)
=13–15(d)	0.00617 (0.0596)	0.0302 (0.0610)	0.0513 (0.0679)	0.146*** (0.0352)	0.125** (0.0424)	0.127* (0.0511)
=16 or more(d)	−0.0560 (0.0744)	−0.0905 (0.0790)	−0.137 (0.102)	0.0241 (0.0550)	−0.0364 (0.0682)	−0.0497 (0.0860)
Mom married in 1968	−0.0485 (0.0607)	−0.0362 (0.0588)	0.0357 (0.0849)	−0.0145 (0.0567)	−0.00132 (0.0592)	−0.122* (0.0531)
Age squared	−0.000436 (0.00075)	−0.000205 (0.000862)	−0.00254 (0.00291)	−0.000294 (0.000698)	−0.000451 (0.00078)	−0.00201 (0.00256)
Black(d)	−0.0141 (0.0715)	−0.0651 (0.0743)	−0.00805 (0.0732)	−0.00340 (0.0552)	−0.0529 (0.0579)	−0.0331 (0.0708)
Hispanic(d)	−0.317 (0.202)	−0.380* (0.257)	−0.425* (0.211)			
pseudo R-sq	0.014	0.020	0.045	0.018	0.024	0.028
Observations	709	728	449	728	712	465

Notes: Marginal effects shown in table. Dummies (d) shows discrete change of dummy variable from 0 to 1. Robust Standard errors in parentheses.

* significant at 10%;

** significant at 5%;

*** significant at 1% Sample of PSID cohort aged 0–16 in 1968 with known mother. Sample of respondents present during 1991. Hispanic cases dropped for females because all SRC Hispanic females in this subsample exit. Variable definitions: Has bad health in 86–91 is ever had fair or poor health during 1986–91; individual education in 1991 when respondent is at least age 24; labor income averaged over ages 25–34. Income, labor income deflated to 2001 dollars.

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

Table 15

Probits for Being Sample Respondent, SRC

	SRC Males with Brothers			SRC Females with Sisters		
Has bad health in 86–91	−0.0538 (0.159)			−0.142 (0.143)		
Individual Education		0.0246* (0.0129)			0.00988 (0.0147)	
Log Labor Income Averaged, age 24–35			0.0222 (0.0442)			−0.00768 (0.0253)
Low birth weight(d)	0.0337 (0.118)	−0.0124 (0.0964)	−0.00864 (0.138)	0.0445 (0.0734)	0.000361 (0.0828)	0.0894 (0.0809)
Log Family income Averaged, age 0–16	0.0623 (0.0659)	0.0448 (0.0511)	0.135 (0.0897)	0.0558 (0.0529)	0.0404 (0.0566)	0.0968 (0.0772)
Age	0.00247 (0.0796)	0.152 (0.172)	0.0905 (0.282)	−0.0342 (0.0729)	−0.0331 (0.0862)	0.197 (0.253)
Birth order number	0.0362* (0.0174)	0.0508** (0.0139)	0.0439* (0.0204)	0.00533 (0.0182)	0.0114 (0.0192)	0.00230 (0.0211)
Mother's Education (<12 omitted)						
=12 years(d)	−0.0342 (0.0600)	−0.0536 (0.0609)	−0.0423 (0.0755)	0.0417 (0.0553)	0.0406 (0.0588)	−0.0166 (0.0759)
=13–15(d)	−0.00288 (0.103)	−0.0388 (0.113)	0.00377 (0.137)	0.222*** (0.0581)	0.212** (0.0666)	0.208* (0.0871)
=16 or more(d)	−0.185 (0.118)	−0.217* (0.122)	−0.264* (0.154)	0.142 (0.0757)	0.128 (0.0864)	0.0695 (0.118)
Mom married 1968	−0.0623 (0.0889)	−0.0472 (0.0914)	0.0355 (0.124)	−0.143* (0.0712)	−0.120 (0.0788)	
Age squared	0.0000242 (0.00126)	0.000589 (0.00151)	−0.00157 (0.00473)	0.000461 (0.00116)	0.000453 (0.00136)	−0.00357 (0.00429)
Black(d)	0.0121 (0.120)	−0.0271 (0.134)	−0.0222 (0.160)	0.0232 (0.0776)	0.0269 (0.0782)	−0.00179 (0.105)
Hispanic(d)	−0.207 (0.256)	−0.275 (0.211)	−0.457 (0.249)			
pseudo R-sq	0.020	0.029	0.048	0.040	0.035	0.039
Observations	380	365	237	411	388	262

Notes: Marginal effects shown in table. Dummies (d) shows discrete change of dummy variable from 0 to 1. Robust Standard errors in parentheses.

* significant at 10%;

** significant at 5%;

*** significant at 1% Sample of PSID cohort aged 0–16 in 1968 with known mother. Sample of respondents present during 1991. Hispanic cases dropped for females because all SRC Hispanic females in this subsample exit. Variable definitions: Has bad health in 86–91 is ever had fair or poor health during 1986–91; individual education in 1991 when respondent is at least age 24; labor income averaged over ages 25–34. Income, labor income deflated to 2001 dollars.

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

Table 16

Probit for Retention of Sibling Pairs, SRC Males

	Levels (1)	Levels (2)	Levels (3)	With Differences (4)	With Differences (5)	With Differences (6)
Has bad health 86–91 (older sibling)	−0.236 (0.159)			−0.188 (0.251)		
Log Labor income, Age 25–34 older sibling		0.105 * (0.0432)			0.0927 * (0.0464)	
High School grad (older sibling)			−0.00357 (0.0804)			0.00744 (0.0930)
Health, Sibling difference				−0.102 (0.197)		
Log Labor Income, Sibling difference					−0.108 * (0.0518)	
High School Grad sibling difference						−0.0428 (0.0893)
Low Birth Weight	0.0156 (0.139)	0.0148 (0.125)	−0.0380 (0.124)	0.181 (0.138)	0.0948 (0.139)	0.166 (0.125)
Log Family income 0 to 16	0.187 * (0.0728)	0.141 (0.0755)	0.179 * (0.0737)	0.142 (0.0783)	0.0851 (0.0801)	0.116 (0.0791)
Low Birth weight, sibling difference				−0.00942 (0.0379)	−0.00884 (0.0319)	−0.0162 (0.0349)
Log Family income 0 to 16, Sibling difference				−0.238 (0.215)	−0.255 (0.209)	−0.367 (0.209)
Mother's Education (<12 years omitted)						
=12 years	−0.0477 (0.0637)	−0.0225 (0.0628)	−0.0139 (0.0620)	−0.0880 (0.0684)	−0.0368 (0.0661)	−0.0589 (0.0686)
=13–15 years	−0.0372 (0.123)	−0.0394 (0.121)	−0.0440 (0.120)	−0.0415 (0.137)	−0.0460 (0.129)	−0.0456 (0.131)
=16 years or more	−0.135 (0.123)	−0.186 (0.119)	−0.155 (0.120)	−0.206 (0.139)	−0.232 (0.126)	−0.230 (0.128)
Pseudo R-sq	0.040	0.043	0.030	0.052	0.060	0.045
Observations	391	408	408	310	363	344

Notes: Marginal effects shown in table; for dummies shows discrete change of dummy variable from 0 to 1. Robust standard errors in parentheses.

* p<0.05,

** p<0.01,

*** p<0.001. Models also include age and age squared of older sib, birth order, whether mother married in 1968, Black and Hispanic indicators. Sample of sibling pairs present in 1991 interview. Both siblings aged 16 or less in 1968. Has bad health is ever had bad health in 1986 to 1991, Individual education in 1991, Labor income average over ages 25 to 34. Income, labor income deflated to 2001 dollars. Sibling differences are in absolute values.

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

Table 17

Probit for Retention of Sibling Pairs, SRC Females

	Levels (1)	Levels (2)	Levels (3)	With Differences (4)	With Differences (5)	With Differences (6)
Has bad health 86–91 (older sibling)	−0.0152 (0.121)			−0.0389 (0.237)		
Log Labor income, Age 25–34 older sibling		−0.00306 (0.0207)			−0.0272 (0.0269)	
High School grad (older sibling)			0.124* (0.0738)			0.214* (0.105)
Health, Sibling difference				−0.0365 (0.211)		
Labor Income, Sibling difference					−0.0407 (0.0275)	
High School Grad sibling difference						0.0221 (0.0958)
Low Birth Weight	0.114 (0.0783)	0.0774 (0.0796)	0.0266 (0.0871)	0.0798 (0.105)	−0.0324 (0.108)	−0.160 (0.136)
Log Family income 0 to 16	.0594 (0.0591)	0.0739 (0.0565)	0.0567 (0.0574)	0.0266 (0.0635)	0.0593 (0.0587)	0.0275 (0.0616)
Low Birth Weight, Sibling difference				0.0370 (0.0271)	0.0480 (0.0270)	0.0710* (0.0310)
Log Family Income 0 to 16 Sibling difference				0.179 (0.171)	0.273 (0.166)	0.336 (0.178)
Mother's Education (<12 years omitted)						
=12 years	0.00891 (0.0611)	0.0282 (0.0620)	0.00552 (0.0631)	−0.00793 (0.0646)	−0.00207 (0.0639)	−0.0374 (0.0679)
=13–15 years	0.171* (0.0841)	0.156 (0.0870)	0.141 (0.0901)	0.207** (0.0799)	0.208* (0.0808)	0.126 (0.0963)
=16 or more years	0.288*** (0.0792)	0.228* (0.0915)	0.238** (0.0909)	0.303*** (0.0684)	0.229** (0.0824)	0.217* (0.0870)
pseudo R-sq	0.053	0.039	0.049	0.062	0.060	0.067
Observations	422	420	421	356	372	335

Notes: Marginal effects shown in table; for dummies shows discrete change of dummy variable from 0 to 1. Robust standard errors in parentheses.

* p<0.05,

** p<0.01,

*** p<0.001. Models also include age and age squared of older sib, birth order, whether mother married in 1968, Black. Hispanic cases dropped because all SRC Hispanic females exit. Sample of sibling pairs present in 1991 interview. Both siblings aged 16 or less in 1968. Has bad health is ever had bad health in 1986 to 1991, Individual education in 1991, Labor income average over ages 25 to 34. Income, labor income deflated to 2001 dollars. Sibling differences are in absolute values.

B E J Econom Anal Policy. Author manuscript; available in PMC 2012 February 23.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Table 18

Summary of Tests for Significant Differences in Coefficients Between Complete 1991 and Retained 2007 Samples

		Males	Females	Source
Sibling correlations	Testwise α =.1	0 (3)	0 (3)	Tables 5B, 5C
	Bonferroni α/C	0(3)	0(3)	
Father Son Earnings Correlations	Testwise α =.1	0 (6)		Table 11
	Bonferroni α/C	0 (6)		
Individual Outcome Regressions	Testwise α =.1	1 (6): Labor	0 (6)	Table 12
	Bonferroni α/C	0 (6)	0(6)	
Fixed Effect Outcome Regressions	Testwise α =.1	0 (6): Labor	3(6): Health, Labor Education	Table 13A, 13B
	Bonferroni α/C	0 (6)	0 (6)	
Individual Retention Probits	Testwise α =.1	2 (3): Education, Labor	0 (3)	Table 14
	Bonferroni α/C	1 (3): Labor	0 (3)	
Sibling Pair Retention Probits (w/differences)	Testwise α =.1	2 (2): Labor	1(2): Education	Table 16, 17
	Bonferroni α/C	0(2)	0(2)	

Notes: Testwise uses significance level applicable to each test separately with α = 1. Bonferroni uses significance level α/C where C is the number of tests in the grouping. The entries show the number of coefficients significant at 10 percent level based on the two criteria, with the number of potential tests shown in parentheses after the count. Example: 1(6) means one of the six coefficients under consideration was significant at the 10 percent level. For significant coefficients, the label of the outcome variable is listed.