Supplement

The code to create analytic data sets, models and plots is available at: https://github.com/sdaza/dissertation/tree/master/ch03. Some of the variables in the paper are restricted and obtained under special contractual arrangements to protect the anonymity of respondents. These data are not available from the authors. Those interested in obtaining PSID restricted data should contact PSIDHelp@isr.umich.edu. Those interested in the NLSY restricted data, visit www.bls.gov/nls/geocodeapp.htm.

Data

Our data result from combining different sources. The first is the Health Inequality Project Data (HIPD) created by Chetty and colleagues (Chetty et al., 2016). Those data – the result of linking 1.4 billion tax records to Social Security Administration records – contain information on income for the period 1999-2014 by U.S. counties and commuting zones.[1] The HIPD also includes statistics of the income distribution and two indicators of income mobility derived from measures of the association between income of children born between 1980 and 1982 and their parents' income.[2] First, we use the index of relative mobility (IRM or *rank-rank slope*) at the county level that is the slope of a regression model between children's national income rank – within a birth cohort – and their parents' national income rank.[3] For the relative income mobility indicator larger values correspond to lower income mobility (i.e., higher rank-rank slope between parents' and child's income). We also use an absolute upward mobility score or "the mean rank (in the national income distribution) of children whose parents are at the 25$^{\text{th}}$ percentile of the national parent income distribution" (Chetty et al., 2014, p. 7).[4] Absolute upward income mobility ranges from 0 to 1, and higher values correspond to larger income mobility. To facilitate interpretation, we multiply the upward mobility score by -1 so that the meaning and expected association of relative and absolute income mobility with health are the same. Finally, we use the Gini coefficient as an indicator of income inequality.

The second database is the National Longitudinal Survey of Youth 1997 (NLSY97), a nationally representative sample of 8,984 American youth born between 1980 and 1984. Surveys were conducted annually, beginning in 1997 when the youth were between 12 and 18 years of age. In the first round, both the eligible youth and one of their parents were administrated personal interviews. The restricted NLSY97 geocoded data file contains information on the geographic residence of each respondent since age 12, allowing us to merge it with Chetty's county level income mobility measures. Importantly, the NLSY 97 sample matches the cohorts of the core sample used by Chetty et al. (2014) (1980-1982), so we can align the timing of early exposure to the place's income mobility. This under the assumption that the income mobility of this cohort measures the *socioeconomic mobility regime* to which this generation was exposed early in life and that may affect their health later. After merging the two databases, we kept 8,810 NLSY97 respondents. Only 174 respondents (2%) were removed from the analytic sample because income mobility information did not match the NLSY97 data. The total number of counties matched was 1607. Figure S1 shows counties included in the NLSY97 sample by income mobility, inequality, and the log of county's population. Figure S1 shows that counties included in the sample have a larger population, less variability and extreme values in the income mobility measures than the counties excluded. The distribution of income inequality is much more symmetric by population size. Although our analyses adjust for sampling weights and county's population size, we note our sample differs from the county composition of previous aggregate and individual level studies that have a higher county coverage. The potential consequences of this coverage difference for our results are discussed later in the *Conclusion and discussion* section of the paper.

The third database is the Panel Study of Income Dynamics (PSID), a nationally representative sample of U.S. men, women, children, and their families followed for more than 40 years. The PSID began interviewing a sample of about 5,000 families in 1968 and were re-interviewed each year through 1997 when the data collection became biennial. Similarly to the NLSY, restricted geographic data allow us to merge individual records with county income mobility measures. Unlike the NLSY 97, the PSID data permit us to estimate the effect of exposure to contextual mobility from birth to age 20. However, we lose statistical power because the number of respondents who match the Chetty et al. (2014)'s cohort is smaller. For instance, between 1975 and 1985, the PSID panel had 4,771 newborns.[5] Of these, 2,358 were the *reference person* or *spouse/partner* of the household at any time during their participation in the panel.[6] Although that cohort does not match exactly the cohort used by Chetty et al. (2014), it offers a reasonable approximation to the mobility regime exposure of that generation, provided income mobility does not change dramatically before 1980-82. After merging PSID and HIPD databases, we obtained 2,273 respondents.[7]. Only 85 respondents (4%) were removed because income mobility information did not match the PSID data. The total of counties matched was 1120, and the distribution of counties by income mobility, inequality, and population size looks similar to the NLSY97.[8] Even though the PSID analytical sample is considerably smaller, we think it is worth to estimate the effects of exposure from birth to age 20 years, and compare those results with the NLSY97.

Using different longitudinal data sources provides a broader picture to examine our research questions. It also offers a more precise definition of exposure to contextual income mobility as both of studies track respondents' county of residence during early life and over a relatively long period of time.

Sample characteristics

Tables S5 and S6 show descriptive statistics of our analytical samples and the proportion of missing data by variable. These tables provide insights about differences regarding design, composition, and length of exposure in NLSY97 and PSID samples. For instance, among NLSY97 respondents, the first interview was, on average, at age 14 (min 12, max 18), while the last interview was at age 33. This contrasts with the PSID sample whose respondents entered the study since they were born, and had their last interview at age 37 on average (min 30, max 47). Due to these differences in measurement, the number of residential changes is also different between samples. Whereas the proportion of NLSY97 respondents who moved to another county during the observation period was 27%, 44% of PSID interviewees have changed their residential county in 20 years.

In addition to these differences, Tables S5 and S6 show the PSID sample has slightly more White respondents than the NLSY97 (59% versus 52%). However, on average, the PSID respondents seem to have lived in counties with a higher proportion of African-Americans (19% versus 15% among NLSY97 respondents). The nature of some covariates also differ. For example, while parents' education is time-invariant in NLSY97 (only measured at the baseline), the PSID recorded that variable over time (i.e., time-variant).

The outcome variables – measured at the end of the follow-up period – show relatively similar values in both samples, except for current smoking. The self-reported health scale (1-5) is around 3.5 points in both samples, while the BMI ranges between 28.7 and 28.9, and the proportion of respondents currently smoking is higher in the NLSY97 (30%) than in the PSID (20%). The remaining outcome variables (depressive symptoms and smoking intensity) are not strictly comparable. The proportion of missing data in the outcome variables ranges between 20% and 31%, and reaches its maximum in BMI (31% in the PSID, and 24% in the NLSY97).

For the rest of the covariates, missing data are considerably lower in the PSID than in the NLSY97. While the PSID's highest proportion of missing cases is observed in the variable *weight of the respondent when was born* (12%), the NLSY97 has considerably higher levels of missing data, especially in time-variant variables such as household income (69%), family size (35%), and parents' working status (32%). This pattern is due, in part, to the design of the NLSY97 study. As the observation window did not always start at age 12, when the first interview was after age 12, no information was collected on several covariates between age 12 and the age of the first interview. In those cases, we had to use retrospective parents' reports to determine where the respondents lived when they were 12 years old and imputed missing covariates during that period.

We note that, although we compare the results from these two datasets, the descriptive tables S5 and S6 show relevant differences in design, composition, and exposure in the NLSY97 and PSID analytical samples that need be considered when interpreting our findings.

Table S5: NLSY97 descriptive statistics of covariates and outcomes

	Mean	SD	Min	Max	% Missing	Valid observations
Time-invariant covariates						
Male	0.51	0.50	0.00	1.00	0.00	8810
Age first interview	14.35	1.49	12.00	18.00	0.00	8810
Age last interview	32.88	1.45	30.00	36.00	0.00	8810
Race-Ethnicity						
White	0.52	0.50	0.00	1.00	0.00	8810
Black	0.26	0.44	0.00	1.00	0.00	8810
Hispanic	0.21	0.41	0.00	1.00	0.00	8810
Mixed	0.01	0.10	0.00	1.00	0.00	8810
ASVAB Test Score	45.38	29.17	0.00	100.00	0.21	8810
Parent's Education (years)	13.15	3.06	1.00	20.00	0.07	8810
Mother's age at birth of respondent	25.48	5.39	12.00	54.00	0.07	8810
Number of residential moves by age 12	3.17	2.75	1.00	40.00	0.13	8810
Proportion moved to a different county	0.27	0.44	0.00	1.00	0.00	8810
Time-variant covariates						
Family size	4.26	1.65	1.00	17.00	0.35	70480
Respondent living with any parent	0.82	0.39	0.00	1.00	0.29	70480
Parent is working	0.89	0.31	0.00	1.00	0.32	70480
Parent is married	0.65	0.48	0.00	1.00	0.31	70480
Log household income	-0.11	2.42	-10.40	2.98	0.69	70480
County log income	0.74	1.08	-3.07	4.32	0.00	70480
County log population	1.76	1.15	-1.62	4.15	0.00	70480
County proportion Black	0.15	0.16	0.00	0.80	0.00	70480
Cumulative number of county moves	0.15	0.47	0.00	6.00	0.00	70480
Exposure variables						
County rank-rank correlation (original)	0.26	0.07	0.04	0.53	0.00	70480
Quintile county rank-rank correlation (original)	2.88	1.37	1.00	5.00	0.00	70480
Residualized county rank-rank correlation	-0.29	0.76	-3.21	2.50	0.00	70480
Quintile residualized county rank-rank correlation	2.52	1.30	1.00	5.00	0.00	70480
County upward mobility (original)	0.44	0.05	0.33	0.67	0.00	70480
Quintile county upward mobility (original)	2.18	1.16	1.00	5.00	0.00	70480
Residualized county upward mobility	0.30	0.64	-2.66	2.34	0.00	70480
Quintile residualized county upward mobility	3.59	1.33	1.00	5.00	0.00	70480
County Gini coefficient (original)	0.45	0.04	0.34	0.60	0.00	70480
Quintile county Gini coefficient (original)	3.45	1.38	1.00	5.00	0.00	70480
Residualized county Gini coefficient	0.21	0.92	-2.84	4.99	0.00	70480
Quintile Residualized county Gini coefficient	3.39	1.45	1.00	5.00	0.00	70480
Outcomes						
Self-reported health	3.63	1.00	1.00	5.00	0.21	8810
BMI	28.70	5.96	15.00	40.00	0.24	8810
Depressive symptoms	1.83	0.50	1.00	4.00	0.21	8810
Current smoking	0.31	0.46	0.00	1.00	0.22	8810
Days smoked	6.21	11.60	0.00	30.00	0.22	8810

Note: Statistics based on non-imputed data. SD = Standard deviation. Observations correspond to respondents in the case of time-invariant and outcome variables, and person-years (N times exposure) for time-variant variables. Outcomes were measured in 2015.

Table S6: PSID descriptive statistics of covariates and outcomes

	Mean	SD	% Missing	Observations
Time-invariant covariates				
Male	0.46	0.50	0.00	2273
Age last interview	37.03	3.25	0.00	2273
Birth year	1980.23	3.18	0.00	2273
Race-Ethnicity				
White	0.59	0.49	0.00	2273
Black	0.39	0.49	0.00	2273
Other	0.02	0.14	0.00	2273
Weighted less than 55 oz	0.07	0.26	0.12	2273
Mother marital status at birth	0.76	0.43	0.04	2273
Mother's age at birth of respondent	25.18	4.95	0.00	2273
Proportion moved to a different county	0.44	0.50	0.00	2273
Time-variant covariates				
Family size	4.30	1.38	0.06	45460
Respondent living with any parent	0.73	0.45	0.06	45460
Parent's years of education	12.87	2.44	0.07	45460
Parent is working	0.61	0.49	0.06	45460
Parent is married	0.83	0.38	0.06	45460
Log household income	0.09	1.13	0.06	45460
County log income	0.64	1.03	0.00	45460
County log population	1.57	1.14	0.00	45460
County proportion Black	0.19	0.19	0.00	45460
Cumulative number of county moves	0.55	1.07	0.00	45460
Exposure variables				
County rank-rank correlation (original)	0.28	0.07	0.00	45460
Quintile county rank-rank correlation (original)	3.27	1.34	0.00	45460
Residualized county rank-rank correlation	-0.20	0.73	0.00	45460
Quintile residualized county rank-rank correlation	2.63	1.36	0.00	45460
County upward mobility (original)	0.43	0.05	0.00	45460
Quintile county upward mobility (original)	2.12	1.21	0.00	45460
Residualized county upward mobility	0.25	0.59	0.00	45460
Quintile residualized county upward mobility	3.56	1.28	0.00	45460
County Gini coefficient (original)	0.45	0.04	0.00	45460
Quintile county Gini coefficient (original)	3.46	1.39	0.00	45460
Residualized county Gini coefficient	0.11	0.83	0.00	45460
Quintile Residualized county Gini coefficient	3.23	1.40	0.00	45460
Outcomes				
Self-reported health	3.53	0.99	0.23	2273
BMI	28.91	5.90	0.31	2273
Depressive symptoms	1.63	0.67	0.23	2273
Current smoking	0.20	0.40	0.23	2273
Number of cigarettes	2.15	5.35	0.23	2273

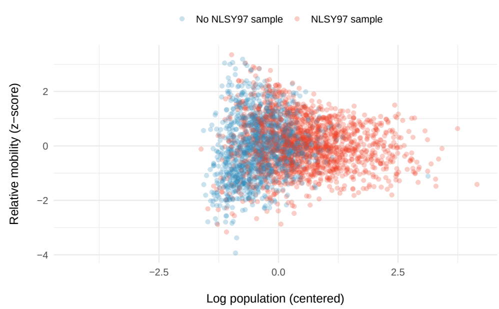
Note: Statistics based on non-imputed data. SD = Standard deviation. Observations correspond to respondents in the case of time-invariant and outcome variables, and person-years (N times exposure) for time-variant variables. Outcomes were measured in 2017.

County coverage

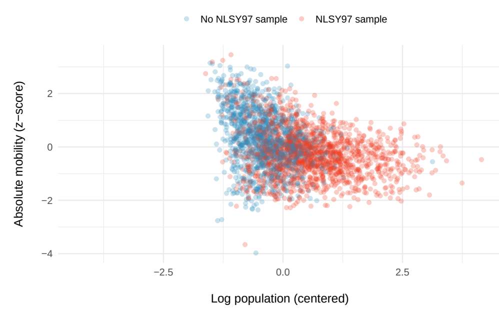
Figure S1 displays counties by the log of population and measures of income mobility and inequality, highlighting in red the counties included in the NLSY sample.

Figure S1: County income mobility and inequality over population
by NLSY97 sample coverage

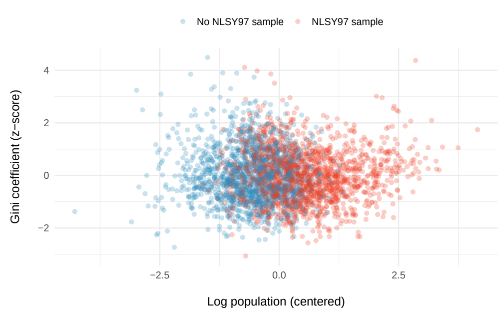
(a) NLSY97 Relative mobility



(b) NLSY97 Absolute mobility



(c) NLSY97 Inequality



This figure provides insights on county coverage of the NLSY97 individual sample, and the relationship between the size of counties and the values of income mobility and inequality. Due to disclousure rules for restricted data, we cannot publish the same plots for the PSID sample. However, the patterns displayed in Figure S1 are similar to what we observed in the PSID sample.

Analytical Strategy

This paper aims to estimate the effect of average exposure to county income mobility during childhood and adolescence on health outcomes such as smoking, BMI, self-reported health, and mental health during young adulthood. The key independent variable is the average income mobility exposure between ages 12 and 20 in the case of the NLSY97, and ages 1 to 20 for the PSID. Outcomes, in contrast, were measured during the last NLSY97 and PSID waves when respondents were in their thirties or forties. As a benchmark, we used both relative and absolute income mobility and estimated the effect of average county income inequality exposure (i.e., Gini coefficient), to compare the magnitude and direction of the associations. We used residualized income mobility and inequality scores from a county-level regression model that adjusted for characteristics such as population size, proportion of African-Americans, average household income, and income inequality (or income mobility). For completeness, we show the results with non-residualized exposure treatments .

We modeled the health outcomes as a function of duration-weighted exposures to different levels of county mobility regimes. By using inverse probability of treatment weighting (IPT), we emulated a counterfactual scenario in which we compared children with the same combination of *observed* covariate values during the exposure time, who did not select systematically into different county mobility regimes. Thus, we adjusted for confounding by time-varying covariates that might be affected by past treatment (Hernán et al., 2002; Hernán & Robins, 2006; Hernán et al., 2000; Wodtke et al., 2011), and generated a pseudo-population in which treatment was no longer confounded with measured covariates. Weights balance treatment assignment across prior confounders and give more or less weight to children with covariates histories that are under-represented (or over-represented) in their current treatment group. To reduce the variability of weights, we used stabilized IPT weights (Hernán et al., 2000; van der Wal & Geskus, 2011). As the estimation of stabilized weights includes time-invariant covariates in the numerator and denominator, final outcomes models need to condition on time-invariant covariates to obtain unbiased estimates of the treatment.

As a sensitivity analysis, and because IPT weights using a continuous treatment are more sensitive to misspecification and outliers (Naimi et al., 2014; Thoemmes & Ong, 2016), we estimated weights for both continuous and categorical scores of income mobility and inequality. While we used linear regression in the first case, we ran ordinal logistic regressions to estimate the probability of exposure to county income mobility quintiles. Finally, following the strategy suggested by Dugoff et al. (2014), we included the sampling weights when computing IPT weights, multiplied them, and considered survey design variables (i.e., strata, clusters) and compound weights when estimating exposure models.

NLSY97

The NLSY97 has information of respondents' location (county) since age 12, so the exposure to county income mobility between ages 12 and 20 can be defined as:

$$\frac{\sum_{i=12}^{20} \text{county income mobility}_i}{8}$$

We employed several covariates to adjust for potential confounding of county income mobility effects on health outcomes. Time-invariant covariates include race, gender, parents' education (years), age by the end of the study (categorical variable), the number of residential moves by age 12, the Armed Services Vocational Aptitude Battery (CAT-ASVAB) score, and mother's age at birth. Time-variants covariates, in turn, are inflation-adjusted family income (log), family size, the cumulative number of county changes, whether parents are employed and married, self-report health status, the number of days smoked in the last month, and BMI.[9] Table S5 shows descriptive statistics of the variables included in our models.

To estimate stabilized IPT weights in Time 1 (Age 12), we employed only time-invariant covariates. From Time 2 to 8, we used both time-invariant, baseline, and lagged time-variant covariates so that weights for later time points included all previous variables.[10] Table S7 shows descriptive statistics for the stabilized IPT weights for both continuous and categorical exposure treatments.

We used multiple imputation with multilevel models to address both item-specific non-response and attrition.[11] By design, respondents interviewed for the first time after their 12$^{\text{th}}$ birthday do not have information between age 12 and the age of the first interview. About 32% of the NLSY97 respondents had full exposure information (i.e., eight interviews from age 12 to 20). On average, respondents reported 6.5 years (out of 8), and only 7% of the sample participated in four or fewer years (e.g., older interviewees). When the county of residence was missing over the follow-up period, we imputed lost counties using most recent or earliest county of residence based on the evidence that most people do not change their county of residence often.[12] Matched NLSY97 counties with the HIPD data cover, on average, 5.9 out of 8 years of exposure, and only 4% of the sample have less than three years. At the end, we imputed missing records by creating 20 multiple imputed data-sets.

We implemented different outcome models depending on the nature of the dependent variable. We estimated ordinal logistic regression for the effect of income mobility and inequality on *self-reported health status* (poor, fair, good, very good, excellent), Generalized linear models (GLM) for *BMI* and *depression symptoms*, logistic regression in the case *current smoking status*, and quasi-Poisson models (also called over-dispersion with quasi-likelihood) for the *number of days smoking in the last month*. Outcome models adjusted only for baseline and time-invariant covariates and took into account sampling design variables (strata, clusters) and weights.

PSID

The PSID sample includes newborns. Thus, we could define the average exposure to county income mobility from age 1 to 20 as:

$$\frac{\sum_{i=1}^{20} \text{county income mobility}_i}{20}$$

We included a relatively similar set of covariates as the NLSY97 sample. Time-invariant covariates involved race, gender, age by the end of the study, mother's age and marital status at birth, and weighed less than 55 pounds at birth. In turn, time-variants covariates were inflation-adjusted family income (log), family size, the cumulative number of county changes, head of household education, whether the head was employed, married, and owns the house where that family was living. Outcomes included self-report health status, BMI, depression, current smoking, and number of cigarettes smoked during the last month.[13] Unlike the NLSY97, PSID outcome variables were not measured systematically during the exposure period. Thus, we decided not to use outcomes as time-variant predictors when estimating IPT weights. Table S6 shows descriptive statistics of the variables used.

We followed the same procedure described in the previous section to estimate IPW weights. First, we computed stabilized IPT weights for Time 1 by including only time-invariant covariates. Then, we created weights from Time 2 to 20 using time-invariant and lagged time-variant covariates so that weights for subsequent time points include all previous variables. Table S8 shows descriptive statistics of the PSID stabilized IPT weights

Again, we used multiple imputation with multilevel models (20 multiple imputed data-sets). When the county of residence was missing during the exposure period, we employed LOCF and NOCB methods. About 44% of the PSID sample moved to a different county during the exposure time. On average, matched counties with the HIPD data cover 17.3 out of 20 years of exposure, and only 1.4% of the whole sample report less than seven years of exposure.

Lastly, we used different outcome models depending on the nature of the dependent variable: ordinal logistic regression when estimating the effect of income mobility and inequality on *self-reported health status*, Generalized linear models (GLM) for *BMI* and *depression symptoms*, logistic regression in the case *current smoking status*, and quasi-Poisson models for the *number of cigarettes smoked during the last month*. The outcome models adjusted only for baseline and time-invariant covariates, and considered sampling design variables (strata, clusters) and weights.

IPT Weights

Tables S7, S8, S11 and S14 show descriptive statistics of the stabilized IPT weights separated by sample and type of exposure variable: income mobility or income inequality, continuous or categorical, residualized or non-residualized. These descriptives come from IPT weights estimated using 20 different datasets with imputed data, and linear or ordinal logistic regression depending on the nature of the exposure variables (continuous or categorical). Because we adjusted attrition through imputation, we did not compute attrition weights.

IPT weights exhibit desirable properties when observed means are close to one, and they have small variance. Tables S7, S8, S11 and S14 show that all estimated weights are well-behaved and centered around one (ranging from 0.98 to 1.08). We found, however, substantial differences regarding the variability of the IPT weights. First, as expected, the variability of weights was much higher when using a continuous exposure treatment than a categorical one (quintile). Second, very high standard deviations of weights were mostly due to outliers. For instance, Tables S11 and S14 show that the standard deviation of the continuous version of the Gini coefficient and upward mobility are considerably large. However, once weights are truncated at the 1^{th} and 99^{th} percentiles, weights become stable, and standard deviations decrease considerably. Thus, we decided to use truncated weights in order to improve the efficiency of estimates and avoid the disproportionate influence of extreme observations (Hernán & Robins, 2006; Thoemmes & Ong, 2016; van der Wal & Geskus, 2011).

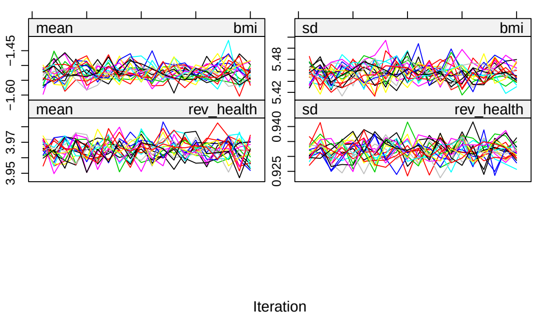
Imputation

We employed multiple imputation for item non-response and attrition. For each exposure variable (e.g., relative and absolute income mobility), we ran multilevel models to impute values for both time-variant and invariant covariates. We produced 20 complete data-sets and pooled the results using Rubin's Rules (van Buuren, 2018).

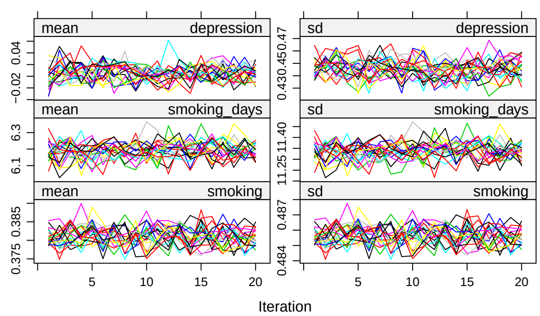
Multiple imputation model specifications are available in our code repository. For each exposure variable, we produced 20 complete datasets per data source (PSID and NLSY9). Different imputed data-sets were created for residualized and non-residualized, continuous and categorical exposure (360 datasets in total). We assessed convergence and feasibility of results using the criteria suggested by van Buuren, 2018. For instance, Figure S2 and S3 show iteration plots of outcomes variables (a and b) and the comparison between observed and imputed distribution of outcomes (d and c). In general, convergence plots looked fine as they mix not systematically after 20 iterations. The distribution of outcomes also seemed reasonable, and no ill behavior of estimates was observed.

Figure S2: NLSY97 Imputation plots with relative mobility as exposure, 20 iterations

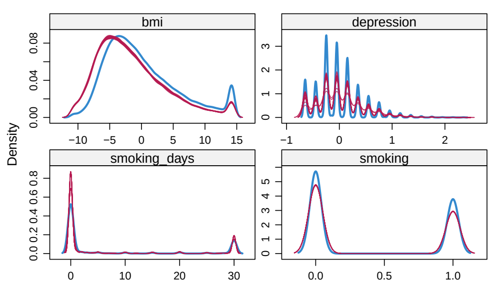
(a) Iterations



(b) Iterations



(c) Imputations vs Observed Values



(d) Imputations vs Observed Values

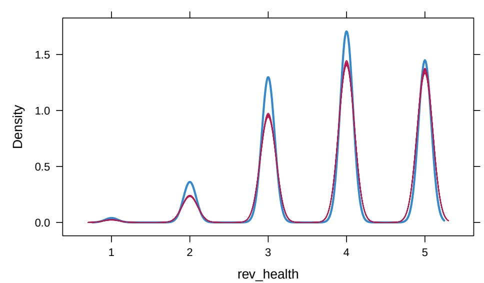
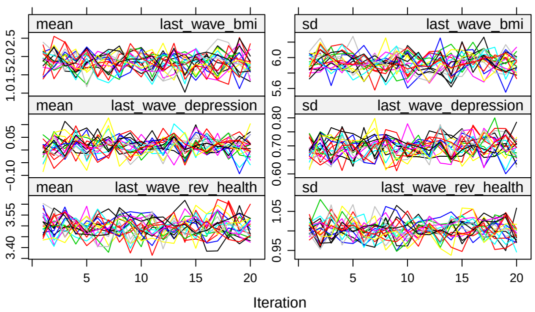
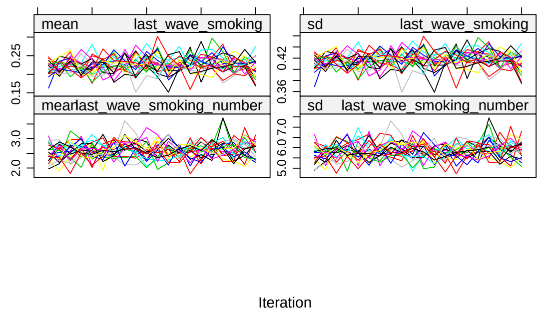


Figure S3: PSID Imputation plots with relative mobility as exposure, 20 iterations

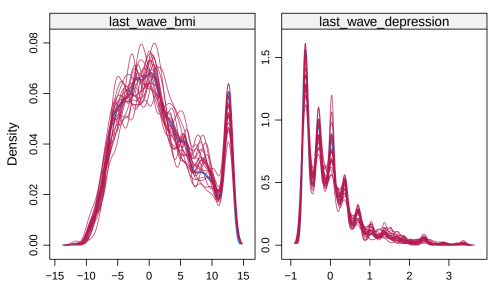
(a) Iterations



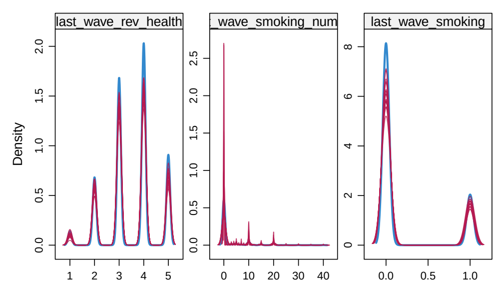
(b) Iterations



(c) Imputations vs Observed Values



(d) Imputations vs Observed Values



IPT weights (residualized models)

Table S7: NLSY97 Stabilized treatment weights (residualized)

			Percentiles			
Weight	Mean	SD	1st	25th	75th	99th
Continuous exposure						
Rank-rank	1.02	0.72	0.28	0.81	1.09	3.07
Upward mobility	1.02	0.48	0.34	0.82	1.09	2.80
Gini	1.02	1.01	0.32	0.84	1.05	3.04
Categorical (quintile) exposure						
Rank-rank	1.00	0.38	0.49	0.83	1.08	2.58
Upward mobility	1.00	0.34	0.54	0.82	1.10	2.27
Gini	1.00	0.41	0.45	0.80	1.10	2.56

Analyses based on exposure from 12 to 20 years old. Statistics based on 20 multiple imputed datasets.

Table S8: PSID Stabilized treatment weights (residual exposure)

			Percentiles			
Weight	Mean	SD	1st	25th	75th	99th
Continuous exposure						
Rank-rank	1.03	0.78	0.34	0.83	1.10	2.75
Upward mobility	1.03	0.90	0.45	0.87	1.08	2.11
Gini	1.00	0.28	0.43	0.89	1.06	2.00
Categorical (quintile) exposure						
Rank-rank	1.00	0.29	0.55	0.82	1.13	1.90
Upward mobility	1.00	0.25	0.59	0.86	1.10	1.73
Gini	1.00	0.24	0.53	0.88	1.09	1.81

Analyses based on exposure from 1 to 20 years old. Statistics based on 20 multiple imputed datasets.

Non-residualized models

For completeness, we show the results with non-residualized exposure treatments and IPT weight statistical descriptives for both the NLSY97 and PSID.

NLSY97

Table S9: Estimates of average continuous exposure on health indicators, NLSY97

	Health status	BMI	Depression	Smoking	Days smoking last month
Unadjusted models					
Rank-rank	-0.04	0.39^{***}	0.00	0.13^{***}	0.15^{***}
	(0.03)	(0.11)	(0.01)	(0.04)	(0.03)
Upward mobility × -1	-0.04	0.30^{**}	-0.00	-0.03	-0.02
	(0.03)	(0.11)	(0.01)	(0.04)	(0.04)
Gini	-0.02	0.16	-0.01	-0.12^{***}	-0.12^{***}
	(0.03)	(0.09)	(0.01)	(0.03)	(0.03)
Adjusted models					
Rank-rank	-0.03	0.28^{*}	0.02	0.11^{*}	0.12^{***}
	(0.03)	(0.13)	(0.01)	(0.05)	(0.04)
Upward mobility × -1	0.04	-0.04	0.01	0.06	0.04
	(0.05)	(0.13)	(0.01)	(0.05)	(0.04)
Gini	0.04	-0.13	-0.00	-0.05	-0.08^{*}
	(0.04)	(0.12)	(0.01)	(0.05)	(0.03)
Individuals	8810	8810	8810	8810	8810

Each coefficient represents a model. Coefficients and standard errors are combined estimates from 20 multiple imputed datasets. Analyses based on exposure from 12 to 20 years old. We estimate different models depending on the outcome: Ordinal regression (self-reported health), General linear model (BMI, depression), Logistic regression (smoking), Quasi-Poisson regression (days smoking last month). $^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Table S10: Estimates of average categorical (quintile) exposure on health indicators, NLSY97

	Health status	BMI	Depression	Smoking	Days smoking last month
Unadjusted models					
Rank-rank	-0.02	0.25^{***}	0.00	0.08^{***}	0.10^{***}
	(0.02)	(0.06)	(0.01)	(0.02)	(0.02)
Upward mobility × -1	0.03	-0.23^{***}	0.00	0.02	0.02
	(0.02)	(0.06)	(0.01)	(0.03)	(0.02)
Gini	-0.01	0.11	-0.01	-0.10^{***}	-0.10^{***}
	(0.02)	(0.07)	(0.00)	(0.02)	(0.02)
Adjusted models					
Rank-rank	-0.01	0.21^{**}	0.01	0.09^{***}	0.08^{***}
	(0.02)	(0.08)	(0.01)	(0.03)	(0.02)
Upward mobility × -1	-0.01	-0.07	-0.01	-0.02	-0.01
	(0.03)	(0.10)	(0.01)	(0.03)	(0.03)
Gini	0.03	-0.07	-0.00	-0.07^{*}	-0.07^{***}
	(0.03)	(0.10)	(0.01)	(0.03)	(0.02)
Individuals	8810	8810	8810	8810	8810

Each coefficient represents a model. Coefficients and standard errors are combined estimates from 20 multiple imputed datasets. Analyses based on exposure from 12 to 20 years old. We estimate different models depending on the outcome: Ordinal regression (self-reported health), General linear model (BMI, depression), Logistic regression (smoking), Quasi-Poisson regression (days smoking last month). $^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Table S11: NLSY97 Stabilized treatment weights

			Percentiles			
Weight	Mean	SD	1st	25th	75th	99th
Continuous exposure						
Rank-rank	1.01	0.91	0.35	0.71	1.05	3.57
Upward mobility	0.99	0.63	0.34	0.74	1.05	3.30
Gini	1.05	1.78	0.29	0.70	1.08	4.29
Categorical (quintile) exposure						
Rank-rank	1.00	0.56	0.39	0.68	1.09	3.51
Upward mobility	1.00	0.55	0.47	0.69	1.08	3.10
Gini	1.00	0.55	0.47	0.67	1.07	3.02

Analyses based on exposure from 12 to 20 years old. Statistics based on 20 multiple imputed datasets.

PSID

Table S12: Estimates of average continuous exposure on health indicators, PSID

	Health status	BMI	Depression	Smoking	Cigarettes smoked
Unadjusted models					
Rank-rank	-0.14^{*}	0.58	0.02	0.13	0.10
	(0.07)	(0.30)	(0.03)	(0.11)	(0.11)
Upward mobility × -1	-0.14^{*}	0.47	0.02	-0.05	-0.03
	(0.07)	(0.28)	(0.03)	(0.14)	(0.12)
Gini	-0.01	0.18	-0.01	-0.12	-0.13^{*}
	(0.08)	(0.22)	(0.02)	(0.08)	(0.07)
Adjusted models					
Rank-rank	-0.06	0.34	0.03	0.10	0.04
	(0.08)	(0.28)	(0.03)	(0.12)	(0.11)
Upward mobility × -1	-0.07	0.58	0.07	-0.05	-0.09
	(0.10)	(0.32)	(0.04)	(0.14)	(0.14)
Gini	-0.02	0.27	-0.02	-0.15	-0.22^{*}
	(0.12)	(0.29)	(0.03)	(0.13)	(0.09)
Individuals	2273	2273	2273	2273	2273

Each coefficient represents a model. Coefficients and standard errors are combined estimates from 20 multiple imputed datasets. Analyses based on exposure from 1 to 20 years old. We estimate different models depending on the outcome: Ordinal regression (self-reported health), General linear model (BMI, depression), Logistic regression (smoking), Quasi-Poisson regression (cigarettes smoked). $^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Table S13: Estimates of average categorical (quintile) exposure on health indicators, PSID

	Health status	BMI	Depression	Smoking	Cigarettes smoked
Unadjusted models					
Rank-rank	−0.07	0.37*	0.01	0.06	0.06
	(0.05)	(0.18)	(0.02)	(0.07)	(0.07)
Upward mobility × -1	0.09	−0.33	−0.01	−0.01	−0.03
	(0.05)	(0.20)	(0.02)	(0.08)	(0.07)
Gini	0.01	0.08	−0.00	−0.08	−0.09
	(0.05)	(0.15)	(0.01)	(0.06)	(0.05)
Adjusted models					
Rank-rank	−0.04	0.16	0.03	0.07	0.05
	(0.05)	(0.17)	(0.02)	(0.07)	(0.07)
Upward mobility × -1	0.07	−0.33	−0.02	−0.15	−0.09
	(0.06)	(0.22)	(0.02)	(0.08)	(0.07)
Gini	0.07	0.06	−0.00	−0.03	−0.07
	(0.06)	(0.18)	(0.02)	(0.07)	(0.06)
Individuals	2273	2273	2273	2273	2273

Each coefficient represents a model. Coefficients and standard errors are combined estimates from 20 multiple imputed datasets. Analyses based on exposure from 1 to 20 years old. We estimate different models depending on the outcome: Ordinal regression (self-reported health), General linear model (BMI, depression), Logistic regression (smoking), Quasi-Poisson regression (cigarettes smoked). $^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Table S14: PSID Stabilized treatment weights

			Percentiles			
Weight	Mean	SD	1st	25th	75th	99th
Continuous exposure						
Rank-rank	1.02	0.81	0.36	0.69	1.07	4.39
Upward mobility	1.04	2.66	0.30	0.57	1.04	4.50
Gini	1.08	1.96	0.24	0.66	1.06	5.52
Categorical (quintile) exposure						
Rank-rank	1.00	0.61	0.41	0.69	1.07	3.40
Upward mobility	0.98	0.70	0.46	0.59	0.98	3.30
Gini	1.01	0.65	0.44	0.64	1.05	3.73

Analyses based on exposure from 1 to 20 years old. Statisticis based on 20 multiple imputed datasets.

Notes

[1]Chetty et al. (2014)'s core sample data include children who (1) have a valid Social Security number or individual taxpayer identification number, (2) were born between 1980 and 1982, and (3) are U.S. citizens as of 2013. There are approximately 10 million children in the core sample.

[2]We use a *permanent-resident* version of income mobility measures, that is, parents who stay in the same counties between 1996-2012. Note that children who grow up in a county may have outmigrated as adults.

[3]Rank-rank slopes have proved to be quite robust across specifications and highly suitable for comparisons across areas (Chetty et al., 2014). *Canonical* measures of relative mobility, such as inter-generational income elasticity (of child income relative to parents' income) tend to be sensitive to changes in inequality across generations. It is also important to note that the rank-rank slopes are not necessarily equivalent to the rank-rank correlation (i.e., Spearman's correlation) within a county. The ranks are computed using the national distribution income not the county distribution. (Chetty et al., 2014), however, argue that both measures are highly correlated at the place level (i.e, counties or commuting zones).

[4]Although at the national level both the relative and absolute measure of mobility provide similar information, when studying small areas, a child's rank in the national income distribution would be an absolute outcome because income in a given area has little impact on the national distribution.

[5]These newborns are PSID *gene* respondents. All 1968 sample members have the PSID *gene*, and they are followed in all subsequent waves across their entire lives, regardless of where they live. All individuals born to or adopted by somebody with the PSID *gene* acquires the gene themselves, and therefore are followed. Respondents who also were the household head or spouse/partner were asked most of our health outcomes overtime and had less missing data.

[6]The outcome variables included in our analysis were mostly asked to *reference persons* and their spouses or partners. That is why we only consider respondents who were a reference person or partner at least once during the observation period.

[7]We also used the PSID *Well being and Daily Life Supplement* 2016 complete missing data of variables such as depression symptoms in the PSID core database.

[8]Due to disclosure rules for restricted PSID data, we cannot show the scatter plot of individuals by county.

[9]The number of cigarettes smoked during the last month was asked only until 2011, that is why, we decided to use the number of days smoked during the last month. Moreover, rounds 4, 6, 8, 10, 12, 14 and 17 of the NLSY97 include a five-item short version of the Mental Health Inventory (MHI-5) to screen for depressive symptoms. Respondents reported the frequency of being nervous, feeling calm and peaceful, feeling downhearted and blue, being happy, and feeling so down in the dumps that nothing could cheer them up using a four-point scale to rate the frequency of their feelings. Because the MHI-5 was only measured in later rounds of the survey, we do not include that scale as a

time-variant covariate.

[10]Details on model specification are available in https://github.com/sdaza/dissertation/tree/master/ch03.

[11]See van Buuren (2018) for an example of selective drop-out correction through multiple imputation.

[12]According to the U.S. Census Bureau migration estimates (Current Population Survey and Annual Social and Economic Supplement 1948-2019), 16% of the U.S. population changed their residence between 1999 and 2000. Of those, 56% remain in the same county (see https://www.census.gov/data/tables/time-series/demo/geographic-mobility/historic.html). In practice, we implemented the Last Observation Carried Forward (LOCF) and Next Observation Carried Backward (NOCB) methods.

[13]The PSID screens mood or anxiety disorder using the Kessler Psychological Distress Scale (K6) in 2001-2003, 2007-2017. The scale includes six items: *During the past 30 days, about how often did you feel nervous, hopeless, restless or fidgety, so depressed that nothing could cheer you up, that everything was an effort, worthless.*

References

Chetty, R., Hendren, N., Kline, P., & Saez, E. (2014). Where is the land of opportunity? the geography of intergenerational mobility in the united states. *The Quarterly Journal of Economics*, *129*(4), 1553–1623. https://doi.org/10.1093/qje/qju022

Chetty, R., Stepner, M., Abraham, S., Lin, S., Scuderi, B., Turner, N., Bergeron, A., & Cutler, D. (2016). The association between income and life expectancy in the united states, 2001-2014. *JAMA*, *315*(16), 1750. https://doi.org/10.1001/jama.2016.4226

Dugoff, E. H., Schuler, M., & Stuart, E. A. (2014). Generalizing observational study results: Applying propensity score methods to complex surveys. *Health Services Research*, *49*(1), 284–303. https://doi.org/10.1111/1475-6773.12090

Hernán, M. A., Brumback, B. A., & Robins, J. M. (2002). Estimating the causal effect of zidovudine on cd4 count with a marginal structural model for repeated measures. *Statistics in medicine*, *21*(12), 1689–1709.

Hernán, M. A., & Robins, J. M. (2006). Estimating causal effects from epidemiological data. *Journal of Epidemiology and Community Health (1979-)*, *60*(7), 578–586.

Hernán, M. Á., Brumback, B., & Robins, J. M. (2000). Marginal structural models to estimate the causal effect of zidovudine on the survival of hiv-positive men. *Epidemiology*, *11*(5), 561–570.

Naimi, A. I., Moodie, E. E., Auger, N., & Kaufman, J. S. (2014). Constructing inverse probability weights for continuous exposures: A comparison of methods. *Epidemiology*, *25*(2), 292–299. https://doi.org/10.1097/EDE.0000000000000053

Thoemmes, F., & Ong, A. D. (2016). A primer on inverse probability of treatment weighting and marginal structural models. *Emerging Adulthood*, *4*(1), 40–59. https://doi.org/10.1177/2167696815621645

van der Wal, W. M., & Geskus, R. B. (2011). Ipw: An r package for inverse probability weighting. *Journal of Statistical Software*, *43*(13), 1–23.

van Buuren, S. (2018). *Flexible imputation of missing data* (Second Edition). Chapman; Hall/CRC.

Wodtke, G. T., Harding, D. J., & Elwert, F. (2011). Neighborhood effects in temporal perspective the impact of long-term exposure to concentrated disadvantage on high school graduation. *American Sociological Review*, *76*(5), 713–736. https://doi.org/10.1177/0003122411420816