The Foerder Institute for Economic Research
at Tel Aviv University

The Eitan Berglas School of Economics

מכון למחקר כלכלי על שם
ד"ר ישעיהו פורדר
על יד אוניברסיטת תל אביב

בית-הספר לכלכלה ע"ש איתן ברגלס

עמותה רשומה

Getting Beneath the Veil of Intergenerational Mobility: Evidence from Three Cities

Oren Danieli, Tanaya Devi and Roland G. Fryer, Jr.

Working Paper No. 10-2021

The Foerder Institute for Economic Research
and
The Sackler Institute of Economic Studies

Getting Beneath the Veil of Intergenerational Mobility: Evidence from Three Cities

Oren Danieli* Tel-Aviv University

Tanaya Devi EO Foundation

Roland G. Fryer, Jr. EO Foundation Harvard University

November 17, 2021

Abstract

We develop a data driven way to design the optimal policy experiment for increasing chances of escaping poverty. We collected data from in-person surveys of almost 1,000 individuals who were reared in poverty in Memphis, Tulsa, and New Orleans, and asked about their childhood health, parental income, home environment as a child, childhood experiences, lifetime traumas, neighborhood safety, a host of psychological skills, beliefs, and current income. Using typical descriptive approaches to motivate an intervention implicitly assumes one can alter individual characteristics in any way the data deem predictive – e.g. sending youth to college to increase future income, regardless of any adverse childhood experiences – even if one rarely observes adults with adverse childhoods going to college in the data. We replace this assumption with four axioms about the cost of altering any combination of individual characteristics. Under these axioms, the optimal experiment replicates the way people escape poverty in real life. We test our method using a case where a data-driven experiment was already run, as well as simulations. We also analyze the robustness of the results if one of the axioms is violated. We find that educational attainment is the most important determinant of mobility. Yet, many other variables – traditionally ignored by economists – are almost equally important predictors: resilience, Big 5 personality skills, grit, self-esteem, the number of adults trusted, trouble with the

*Email: orendanieli@tauex.tau.ac.il. Oren gratefully acknowledge financial support from the Sapir Center for Economic Development, and the Foerder institute for Economic Studies.

police when young, and other adverse childhood experiences. Fathers present in own neighborhood did not matter. This suggest that income-increasing interventions for the poor need to be broader than simply human capital or place-based policies.

1 Introduction

Concerns about poverty and intergenerational mobility are as old as civilization itself.[1] Thousands of years later, poverty and intergenerational mobility are, still, among the most important economic and social issues of our time. In an average OECD country it would take four to five generations for children in the bottom earnings decile to attain the level of mean earnings, but there is significant heterogeneity across countries and between ethnic and racial groups within countries. OECD's annual report on social mobility estimates it will take two generations for children in the bottom decile in Nordic countries to reach the mean; four to six generations across Europe, and five generations in America (OECD, 2018).

Within the US, differences across racial groups are stunning. For some, America is the land of unparalleled opportunity. For others, it is the land of the ineludible poverty trap. Chetty, et al (2018) employ a large new dataset – linking census data covering the U.S population to federal income tax returns from 1989 to 2015, a total of 20 million observations – to estimate intergenerational mobility in America, by race and for granular geographies, providing more precise estimates than Solon (1992).[2] They find that black Americans have lower rates of upward mobility and higher rates of downward mobility compared to white Americans and on par with Native Americans. Blacks in the top income percentile have male children who are as likely to be incarcerated as whites in the 20th percentile.

Gaining a better understanding of intergenerational mobility is of great importance. If, by accident of birth, certain individuals are not able to achieve their full potential then there are important imperfections in the market for talent and making that market more meritocratic may have large social value.

A wide variety of possible correlates of intergenerational mobility have been put forth. These include education and school quality (Aldaz-Caroll and Moran, 2001; McKernan and Ratcliffe, 2005; Haskins, 2008; Baum, Ma and Payea, 2013), neighborhood quality (Keels et al., 2005; Sanbonmatsu et al., 2012; Chetty et al., 2014; Chetty et al., 2016; Chetty and Hendren, 2018), early childhood and adverse childhood experiences (Duncan and Rogers, 1988; Aldaz-Carroll and Moran, 2001; Metzler et al., 2017), family structure and parenting (Duncan and Rodgers, 1988; Aldaz-Carroll and Moran, 2001), and church going

[1] In the oldest written text, "Gilgamesch", there are mentions of famine and descriptions of poverty are in Confucius writings, the Iliad, and the Odyssey. Inequality and mobility were discussed in Ancient Egypt during the rein of King Akhenaten – 80% of the wealth of the belonged to 20% of the population.

[2] Solon (1992) used the PSID dataset and corrected for measurement error to show that the correlation between fathers' earnings and sons' earnings is approximately 0.4. This proved that US was not as highly mobile a society as was previously believed – previous estimates of the correlation were of the order of 0.2.

and other forms of social capital (Freeman, 1986; Sharkey and Torrats-Espinosa, 2015; Western and Pettit, 2010).

In this paper, we attempt to shed new light on the correlates of mobility in America using new data from in-person interviews of approximately one thousand families – all of which self-report growing up poor – in Memphis, Tennessee, New Orleans, Lousianna, and Tulsa, Oklahoma.

Interviews were conducted in a respondent's home or a public place, whichever was preferred, and lasted almost two hours.[3] At the end of the interview, the respondent received a pre-paid $150 Visa gift card. Our final sample consists of 928 respondents for the full interview.

The extensive length of our face-to-face interviews allowed us to collect a wide-ranging set of data: basic demographics, mental health, physical health, parental behavior, home environment, childhood experiences, risky attitudes as a teenager, lifetime traumas, neighborhood safety, and psychological skills such as the Big 5 personality traits, grit, locus of control and resilience. Our main outcome variable is log household income in 2016, though we also present results for individual income, adult mental and physical health, and drug and alcohol use.

The results we obtain from these new data are informative, surprising, and inspire the development of new methods. Using typical descriptive approaches, such as those implemented in Garces, Thomas and Currie (2002), the correlates of intergenerational mobility are education, resilience, mental health before age 16, trouble with police before age 18 and grit. Variables such as childhood abuse, parenting, teenage risky behaviors, trust in adults, or other psychological skills are seemingly not important. Moreover using standard descriptive methods to identify which variables are important, when the number of covariates is high, tend to generate results that are not robust to minimal changes in sample size or variable definitions (Mullainathan and Spiess, 2017).

This inspired us to think about the approaches used in the literature. Designing interventions based on results from prediction methods such as least Squares estimates or popular supervised learning algorithms implicitly assume one can alter individual characteristics in any way the data suggest is optimal – even if one rarely, if ever, observes individuals in the data with those combinations of characteristics. Consider the follow-

[3]To screen potential survey respondents, individuals were first contacted by phone. A respondent was deemed eligible if: (1) they resided in a zipcode in our desired geographies; (2) were at least 18 years of age; (3) self-identified as having "grown up poor" (e.g. answered "Do you consider yourself to have grown up poor?" in the affirmative). This initial conversation lasted, on average, six minutes, and allowed us to collect information on basic demographics, education, and preferred contact information for all eligible respondents. In total, 457,317 phone calls were made and 6,459 were completed.

ing thought experiment. Imagine that graduating from college has large causal effects on future income for everyone, but it's extremely rare that individuals who have endured childhood abuse graduate from college. Just estimating typical descriptive methods – trying to predict income conditional on observables – may tell us to simply get everyone to graduate from college and this will improve their income. But it won't tell us if it's possible!

We relax this assumption and replace it with four axioms on the cost function of an intervention. We assume cost functions are continuous, weakly positive, invariant to representation and additive. The first two axioms are standard and innocuous requirement from a cost function. Invariance to representation guarantees a method that will be applicable in different contexts independent of the specific details of the problem. Additivity nails down a specific functional form.

We find a unique cost function that satisfies these four axioms. We prove that up to scaling, this function is the Kullback-Liebler divergence between the distribution of variables before and after the intervention. It measures how similar are those distributions. Intuitively, cost is higher for interventions that make larger changes in the distribution of child characteristics.

We therefore use the distribution of covariates to understand how our variables relate to one another and estimate the costs of altering any combination of individual characteristics. The joint distribution of the independent variables contains important information that is typically overlooked. In the thought experiment above, this implies that if we observe the lack of childhood abuse and college graduation typically covary, we infer that it is difficult to ensure that victims of childhood abuse graduate from college. And, thus, any income-increasing intervention may want to target both.

Our approach boils down to a simple maximization problem with two key parameters: the distribution of individual characteristics and the distribution of the outcome variable, conditional upon those characteristics. The solution of this problem is a distribution of characteristics that maximizes the predicted outcome variable given this characteristics, at a given budget constraint. While our method cannot tell us what is the intervention that would generate such a distribution, it is still informative about which interventions we should consider. For instance, by knowing that after the optimal intervention psychological characteristics are going to improve dramatically, while other variables related to neighborhood, parental practices or physical health remain similar, could focus our efforts on interventions related to improving non-cognitive skills.

The solution to this maximization problem is simple: interventions should focus on ways in which people typically escape poverty in reality. This solution is unsurprising as

we are looking for improving adult income, using small changes to the current distribution. We show that under linear and normality assumption, our method is equivalent to Pearson correlations. Without any parametric assumptions, we get a simple reweighting procedure (similar to DiNardo et al., 1996), where higher income people receive larger weights.

We believe this method can be used anytime one wants to use observational data to better optimize social experiments designed to increase some desired outcome (e.g. test scores, labor market participation, income in developing countries). While our method does not imply anything on causality, it highlights potential experiments, that if work, could generate very cost-effective policy interventions. While there is surely value in designing experiments based on intuition or qualitative research, we think experiments could improve by relying more on descriptive analysis. Many experiments are already motivated by some descriptive analysis that focuses on a small subset of factors on which data is available. We offer a more systematic approach of collecting a wide data on a large number of potential causal factors, and use our method to understand which intervention is most promising.

When we use our method on our data, we find that the correlates of intergenerational mobility are quite different. Education is still the most important factor in intergenerational mobility. Of the eight other correlates that are significant, however, five of them are psychological skills: resilience, the Big 5, self-esteem, self-control, and grit. The remaining three are whether the respondent was in trouble with the police in their youth, the number of adverse childhood experiences – such as abuse – and the number of adult relationships they trusted during their childhood. The fact that education is the most persistent correlate of mobility is consistent with more than a half century of scholarship in economics. Beginning with Blau and Duncan (1967) and formalized in Becker and Tomes (1979), the model of mobility has been the quantity of skills and their prevailing market price. Our results, along with the burgeoning literature on the importance of non-cognitive skills, suggest a much larger set of skills and experiences are important to produce income. Admittedly, the importance of psychological skills in the production of earnings is not a new idea – Heckman et al. (2006) state that "...for many labor market outcomes, a change in non-cognitive skills...has an effect on behavior comparable to or greater than a corresponding change in cognitive skills" – but its relative importance to human capital and family environment is striking.

Variables that capture place-based heterogeneity such as the neighborhood specific probability of the bottom 25 in top 20 percentile are not significant in our main specifications, and are ranked below psychological variables in all specifications. The explanation

is simple: while some neighborhoods are generating much higher chances of escaping poverty, such neighborhoods are rare, and especially for blacks (Chetty et al., 2018). The more common characteristics of children who escape poverty are the ones we identify in our method.

We explore the robustness of our results across various measures of income, alternative specifications, and alternative measures of adult well-being. Our results are virtually unchanged when using adjusted household income or individual income as outcomes variables. We also receive similar results when we use college graduation as our outcome variable. Alternative specifications, such as allowing for heterogeneity in the distribution of individual characteristics or not adjusting for poverty levels as a child, also yield similar results. Interesting differences do emerge when we use mental illness, or drug and alcohol use as alternative measures of adult well-being. The most important correlates of mental illness include various psychological skills, risky behaviors as a teenager, mental and physical illness as a teenager, education, adverse childhood experiences, family environment, interactions with police, number of adult relationships trusted, neighborhood mobility and an index of neighborhood safety. The latter is consistent with data gleaned from the Moving to Opportunity experiment (Sanbonmatsu et al., 2012).

Similarly, the most important correlates of adult drug and alcohol use are self-control, risky attitudes as a teenager, whether an individual was in trouble with police as a youth, mental health, relationship with parents, neighborhood safety, family environment, and adverse childhood experiences. While education was the most persistent predictor of income, it is a significantly less important predictor of adult mental illness or drug and alcohol use as an adult. Psychological characteristics are also important correlates for marital status, including our measurement of patience which was not significant for our income analysis.

We use three different approaches to test whether our method is actually useful in designing policy experiments. First, we test it in a case where an experiment has already been run. Dobbie & Fryer (2013) collected a wide data on charter school characteristics and linked it to their performance. We use our method on this data and find which variables are most promising for an intervention. An intervention that was targeting a similar set of variables was conducted by Fryer (2014), and was indeed effective in substantially increasing students math test scores.

Second, we analyze the robustness of our results if our axioms our violated. One particular concern is that some variables are harder to change than others, which would violate our axiom of irrelevance of representation. We develop a robustness test to measure how results change when some moments of the distribution are harder to change. We

use it to show that psychological characteristics are still the most important variables (together with years of schooling), even if their cost is 20 times higher than other variables. We also show that we get qualitatively similar results without assuming additivity.

Third, we use examples that we can solve analytically or using simulations to compare our method with existing descriptive methods. We characterize the conditions in which our method is likely to outperform other methods. These are cases where (1) some variables are only operating indirectly through other variables and (2) there are not many side-effect variables that are changing due to interventions in other variables, but do not have any impact on the outcome.

Our analysis has three important caveats. First, the locations are not representative. We chose them because they are representative of a significant share of black poverty in America. Appendix Table 2 compares our survey sample to a comparative sample of National Longitudinal Survey of Youth 1979 (hereafter NLSY) respondents who were between the ages of 14 and 22 *and* below the poverty line in 1979. Strikingly, our sample – taken from these three cities – is statistically similar to the unadjusted nationally representative data on most demographics. The only variables that are statistically different are age and percent Hispanic.

Second, our survey design requires adults to remember their childhood details with a degree of specificity and objectivity that may be implausible.[4] Adding to this complication, our variables designed to assess psychological capital and beliefs – grit, growth mindset, resilience, and so on – are contemporaneous. It is plausible that increasing ones income from childhood poverty causes one to remember childhood experiences in a different light, tell oneself a different narrative or feel more resilient. We had to choose between our current design, which has important shortcomings, and starting a longitudinal dataset similar to the NLSY and waiting at least two decades for the results.[5]

One way to try and understand how this may affect our results is to analyze any individual characteristics in our data that may have been sampled multiple times in a known longitudinal dataset. NLSY, for instance, assessed the locus of control of its respondents in 1979 when individuals were 14 to 22 and again in 2014 when they were older adults. The correlation between locus of control and income in our data is equivalent to correlating 2014 income with 2014 measures of locus of control in the NLSY; both are positive

[4]There is some evidence that individuals are more likely to discuss sensitive topics such as abuse with strangers after the fact and not necessarily contemporaneously (Alaggia, 2010).

[5]Another possibility was to ask parents or friends about the respondent when they were young. This might provide a measure of psychological capital when the respondent was young or a different perspective on sensitive issues such as abuse. This method is similarly problematic and, for a fixed budget, reduces our sample size significantly. We chose a larger sample.

and significant. Importantly, this correlation in the NLSY exists and is of similar magnitude whether one correlates the 2014 measure of locus of control with 2014 income or the 1979 measure of locus of control with 2014 income. Put differently, it seems that there is a strong correlation between measures of locus of control in youth and adult income and that correlation is almost identical if one uses a measure of locus of control assessed in adulthood.

Third, which is less caveat than clarification, our results are correlates of income and other measures of adult well-being which suggest what types of interventions will be most successful – not causal estimates.[6] Without a well-powered field experiment or valid instrument, thorny issues of self-selection, omitted variable bias, and the like may influence our results. Yet, without any descriptive analysis, it is hard to assess which experiment is more likely to generate a meaningful impact. We view our approach and field experiments as strong complements. This paper is an important first step in a larger research agenda whose aim it is to take these correlates and conduct a large scale randomized control trial with a subset of them, aiming to increase intergenerational mobility and overall life outcomes of those who, by circumstance of birth, are more susceptible to continue to be low income in future generations.

With the above caveats and clarifications in mind, our paper makes three contributions. First, we collect new detailed data on individuals who were born into poverty. This data is more comprehensive than previous datasets, including information on sensitive issues such as abuse, relationships with parents and other adults, interactions with police, mental and physical health, learning disabilities, and so on. Second, since Fisher (1925), randomized control trials (RCTs) have grown tremendously in use and importance.[7] The methods developed provide a way to use rich observational data to potentially make those experiments more effective, which could save billions of dollars and alter millions of lives. Third, the results from our new data and new approach, offer an innovative way forward for increasing intergenerational mobility in America.

The paper is organized as follows. The next section details our sample frame and the

[6]Yet, evidence from experiments or natural experiments, suggests that 7 of our identified correlates of income mobility seem to have a causal effect as well . For instance, Angrist and Krueger (1991) demonstrate that compulsory schooling laws increase years of schooling and hence, higher wages. Heckman et al. (2013) show that participants of the HighScope Perry Preschool Program have higher measures of the Big 5 personality traits and increased monthly income at age 27.

[7]The intellectual development of RCTs is varied. Many are theory driven – testing important social scientific theories in the field (e.g. The impact of teacher specialization on student achievement, Fryer 2018). Others seem more resource driven – the federal government spends $565 billion per year on medicaid and we don't know how effective these investments are (Finkelstein et al., 2012). And, many are impact driven – understanding how best to increase student achievement, employment rates, income in third world countries, or reduce crime. Our method applies to the last category.

data collected for our analysis. Section 3 describes our methods and section 4 reports empirical results from combining the new methods and data. Section 5 tests our method ability to inform the design of experiments. Section 6 concludes.

2 A New Survey of Intergenerational Mobility in America

A. Design

In choosing which cities to conduct our survey, we were interested in selecting areas with high levels of poverty, ethnic diversity, and geographic variety. We settled on the general areas of Memphis, TN, Tulsa, OK, and New Orleans, LA. To more precisely define the areas, we started with the Metropolitan Statistical Areas that contained these cities and then selected four counties within them: Shelby County, Tulsa County, and Jefferson and Orleans Parishes. Shelby County has a population of 936,961 with 54.1% black, 6.4% Hispanic, and 35.9% non-Hispanic white. Twenty-one percent of the population currently lives in poverty. Tulsa County has a population of 646,246 with 10.8% black, 12.7% Hispanic, and 60.2% non-Hispanic white. Sixteen percent of the population currently lives in poverty. Since the counties in the New Orleans MSA had smaller populations, we selected two: Jefferson and Orleans Parishes. Jefferson Parish has a population of 439,036 with 27.6% black, 14.9% Hispanic, and 52.5% non-Hispanic white. Sixteen percent of the population lives in poverty. Orleans Parish has a population of 393,292 with 60.1% black, 5.7% Hispanic, and 30.7% non-Hispanic white. Twenty-four percent live in poverty.

B. Sample Selection Method

After determining our sample areas, the next step was to decide how to select our sample within those areas. Previous surveys have often relied on address-based sampling (ABS) to help ensure a representative sample. For example, in the National Longitudinal Survey of Youth 1979, interviewers went to a random sampling of housing units and performed a short screener in person. Although this method is often considered the gold standard for in-person interviews, it is also very expensive. An alternative method of screening is by phone. In order to determine if the samples obtained by phone-based screens and ABS were comparable, we ran a pilot study in Los Angeles County, California, consisting of 643 residents. After comparing the two samples and finding no significant differences in the demographics of the populations surveyed, we selected the phone

screening method as it is much more cost-effective and allowed us to interview more individuals for our final sample. (See Appendix D.3 for a full methods report on our survey design). Abt Associates was responsible for the implementation of both the pilot and full interview.

C. Phone screens

In order to be eligible for the full interview, individuals had to reside in a zip code that was in one of our sampling counties, be at least 18 years old, and self-identify as having grown up poor. Sixty-five percent of the screening frame came from a cell phone screened sample, and thirty-five percent came from a landline screened sample. During the phone screen, we collected information on basic demographics including gender, race, education and contact information for those individuals who were eligible. (See Appendix D.4 for the full text of phone screen). The phone screens lasted an average of 6.3 minutes. In total 6,459 phone screens were completed in our three sampling areas: 1,227 were eligible and agreed to participate in the full interview, 1,390 were eligible but refused further participation and 3,842 were ineligible. (See Table in Appendix D.2.5 for full distribution of respondents by location).

D. Interviews

For subjects who were eligible and agreed to participate, in-person interviews were scheduled. These interviews lasted an average of 104 minutes; approximately 350 questions were asked. The majority of interviews were conducted in individuals' homes, although interviewers were also willing to meet with respondents in public places like coffee shops or the library if the respondent preferred. At the end of the interview, the respondent received a prepaid $150 Visa gift card. During the interviews, we asked questions on a wide variety of topics. We incorporated topics that have been considered potential causes of poverty in the literature and focused on questions and psychological scales that have been developed and validated in previous studies. Below, we describe the general categories of topics we were interested in along with some of the major subsections. [8]

E. Income and Adult Well-Being

[8] The full interview text is available in the Online Appendix D.5.

The main outcome variables used in our analysis are (log) individual income and household income in 2016 in dollars, a series of detailed questions designed to assess mental health, and a set of questions to assess drug and alcohol abuse. Of 928 respondents, 764 were willing and able to answer the income question in an open-ended format. If a respondent said that she did not know or want to answer, we asked her if she would be willing to tell us what range it fell within ($0-$10,000; $10,000-$20,000; $20,000-$30,000; $30,000-$40,000; $40,000-$50,000; $50,000-$75,000; $75,000-$100,000; $100,000-$150,000 or more than $150,000) and then assigned her the midpoint of that range.

Beyond income, we analyze two additional measures of adult well-being: mental illness and drug and alcohol abuse. To assess these measures, we used the industry standard instruments for screening in clinical settings. This includes CAGE-AID: a questionnaire developed to screen for drug use (Basu et al., 2016); GAD 7: a 7-item self-administered questionnaire developed to measure and assess generalized anxiety disorder (Spitzer et al, 2006); and PHQ-9, a 9-question instrument used for screening, diagnosing, monitoring, and measuring the severity of depression (Spitzer et al 1999). These instruments have been used widely in research and clinical practice and generally have been shown to have superior validity when compared to alternative screening questionnaires as well as reliability with independent diagnoses conducted by mental health professionals (Kroenke, Spitzer and Williams, 2001; Lowe et al, 2004; Brown et al, 1998; Brown and Rounds, 1995; Spitzer et al, 2006). See Appendix D.3for details.

F. Other Variables

Basic Demographics

We collected a number of demographic variables in both the phone screen and full interview. This includes gender, race (coded as white, black, Hispanic or other), age, current household members, and highest level of education completed. Our education variable asked individuals to classify themselves as not completing high school, having a high school degree/GED/some college, having a two year Associate's degree, or having at least a Bachelor's degree.

Parental Income

Although we do not contain actual estimates of parental income, we collected variables that proxy for parents' economic conditions. These include participants' responses to questions about whether their families were well off, average or poor financially; whether

they ever moved due to financial difficulties or ever received help due to financial difficulties; if there was a time when the father was unemployed for several months; if the mother ever received welfare; and the frequency with which their families (a) found it difficult to afford child care, (b) fell behind rent or mortgage payments, (c) fell behind gas, electric or phone bill payments, (d) were unable to pay for transportation to get to work or school, (e) were unable to afford medical care, (f) had trouble paying a credit card balance, and (g) had too little money to buy enough food. To ensure that we are targeting people who grew up poor only, we drop participants who say that they had grown up "well off" . This reduces the sample size by 28 respondents.

In our analysis, we aggregated these variables into a "Parental Income Index". We construct a separate parental income index for each outcome variable we use. This index is simply the predicted level of the outcome, conditional on all these variables, using a linear model. We control for this parental income index in all of our specifications, as we want to focus on interventions on children, not their parents.

Early Home Environment

These variables were meant to capture the environment that an individual was raised in through questions about childhood household composition, neighborhood safety, and financial difficulties in their childhood. Questions were mainly drawn from surveys administered as part of the Moving to Opportunity experiment and the Health and Retirement Survey. We were particularly interested in looking at the roles and practices of parents or parental figures and other important relationships in childhood. We relied on the Short Version of the Family Environment Survey, a 27-item inventory designed to measure social-environmental characteristics of the family that was developed by Rudolf Moos and Bernice Moos in 1994. This scale features items such as "Family members had strict ideas about what is right or wrong" and "There were very few rules to follow in our family" and asked respondents to state whether these statements were true or false of the family they lived with between the ages of 5 and 12. Additionally, we adapted questions from the Parent Practices Survey, a 34-item self-reported instrument designed by Dr. Joseph Strayhorn to understand parents' patterns of interaction with their preschool children. In a sample of 200 low-income parents, the scale had good internal consistency and 6-month stability (Strayhorn and Weidman, 1988) and was associated with measures of parents' psychological and social health. Since the original survey was targeted towards parents, we reworded questions to ask respondents their perceptions of their parents' attitudes. For example, "How often does this child do something that gives you pleasure and enjoyment" was rewritten as "How often would your parent say that you did something

that gave him/her pleasure and enjoyment?"

Childhood Traumas and Risky Behaviors

Our main instrument for assessing childhood trauma is the questionnaire from the Adverse Childhood Experience (ACE) study, a 10-item self-reported measure developed to quantify cumulative childhood stress (Felitti et al., 1998). The study, and further follow ups, repeatedly found a relationship between negative later life health outcomes (e.g. alcoholism) and increasing numbers of stressors (e.g. an ACE). ACEs include items such as whether an individual experienced childhood physical or sexual abuse, negligence, witnessed the physical abuse of another household member or whether another household member was mentally ill or addicted to drugs. We investigate whether this relationship continues to be robust in a population of mainly low income individuals and if applied to outcomes outside of the health domain. Additionally, we used a list of lifetime traumas that an individual may have experienced before the age of 18 (such as being in trouble with the police, repeating a year of school or being abandoned by his parents) along with questions on childhood health that were taken from the Psychosocial and Lifestyle Questionnaire of the Health and Retirement Survey. Questions about an individual's experience with drugs and alcohol during childhood were drawn from the 2013 National Youth Risk Behavior Survey. Finally, questions about risky attitudes and behaviors during adolescence were taken from the Moving To Opportunity Child Questionnaire.

Psychological Skills

We chose a broad spectrum of psychological scales, some of which have been tested in large surveys previously (e.g. Rotter's locus of control scale), as well as some that have been developed more recently (e.g. Grit Scale, Dweck Mindset Instrument). Where available, we used abridged scales to limit the length of the survey. Below, we describe a selection of these scales.

The Brief Resilience Scale is a 6-item scale developed by Smith et al. in 2008 to assess the ability to recover from stress. It has been examined in samples of students as well as chronic pain and cardiac patients and found to be reliable and negatively related to mental and physical health symptoms including anxiety and depression. The scale includes items such as "I tend to bounce back quickly after hard times" and asks respondents to state whether they strongly disagree, disagree, neither agree nor disagree, agree, or strongly agree. A meta-analysis of various resilience scales used in the field found that the Brief Resilience Scale was among one of the highest rated scales in terms of construct validity and internal consistency (Windle et al., 2011).

The Brief Self-Control Scale (BSCS) was developed by Tangney, Baumeister and Boone

in 2004 and is used to measure the five domains of self-control: self-discipline, resistance to impulsivity, healthy habits, work ethic and reliability. The scale was found to have good internal consistency and retest reliability. Higher scores on the self-control scale were correlated with lower rates of alcohol abuse, higher grade point averages and better interpersonal skills. The BSCS has been used in over 100 published studies on adolescents and adults to predict numerous behavioral outcomes including high school achievement, job-searching behavior, binge eating and work performance (Lindner et al., 2015; Duckworth and Seligman, 2005; Baay et al., 2015; De Ridder et al., 2012)

Rotter's locus of control scale is a 23 item scale developed by Julian Rotter in 1966 to assess the extent to which an individual feels he can control his circumstances and outcomes. We used an abridged version that relied on four items to mirror the version used in NLSY79. Each item contains two statements (e.g. A. In my case getting what I want has little or nothing to do with luck and B. Many times we might just as well decide what to do by flipping a coin.) and asks respondents to select which statement is closer to their opinion. The four-item version used in NLSY was found to correlate with schooling decisions, employment and wages (Heckman et al., 2006; Darity et al., 1997).

The Dweck Mindset Scale was developed by psychologist Carol Dweck and is used to differentiate between a fixed mindset, in which individuals believe basic qualities like intelligence are fixed traits, and a growth mindset, in which individual believe that their abilities can be developed. A study of high schoolers in Chile found that having a growth mindset strong predicted academic achievement, particularly among low-income students (Claro et al., 2016). Similarly, a longitudinal study of middle school students found that those with a growth mindset outperformed students with a fixed mindset in mathematics two years later (Blackwell et al., 2007). We used an abridged 3-item version that focuses on fixed views of intelligence such as "Your intelligence is something about you that you can't change very much" (Dweck, 1999).

The Rosenberg Self Esteem Index is a 10-item scale developed by Dr. Morris Rosenberg in 1965 to measure both positive and negative feelings a respondent may have about himself. It has been widely used across fields and in large surveys including in the NLSY79. In the NLSY79, higher measures of self-esteem were correlated with future economic success (Heckman et al., 2006). Respondents are asked to indicate their level of agreement with statements such as "On the whole, I am satisfied with myself" and "I feel I do not have much to be proud of."

The original 12-item Grit Scale was developed by Duckworth et al. in 2007 and is meant to measure the ability to sustain effort towards long-term goals. We used the 8-item Short Grit Scale which has been found to correlate with retention in cadets attending West

Point, higher educational attainment and fewer job changes among adults (Duckworth et al., 2007; Duckworth and Quinn, 2009). The 8-item scale contains statements like "New ideas and projects sometimes distract me from previous ones" and "I often set a goal but later choose to pursue a different one" and asks respondents to indicate their agreement with the statement on a five-level Likert scale.

To assess personality traits, we first started with the the International Personality Item Pool, a site with over 3,000 items and 250 scales that are used to measure personality traits (Goldberg, 2006). We selected the 50-item sample questionnaire based on Goldberg's markers for the Big-Five domains of personality: extroversion, agreeableness, conscientiousness, emotional stability, and intellect (Goldberg, 1992). The 50-item Big-Five scale (IPIP) has good internal consistency and related strongly to two other leading personality questionnaires – NEO Five Factor Inventory and Eysenck Personality Questionnaire Short Form (Gow et al., 2005). Respondents are asked to indicate how accurate they think statements such as "I am interested in people" or "I pay attention to details" are in describing themselves.

G. Descriptive Statistics

Our final sample consists of 900 respondents. Appendix Table 4 provides brief sample accounting. We made 458,317 phone calls; 6,459 completed the initial phone screen; 2,617 were eligible for the full study and 3,482 were deemed ineligible. 1,227 agreed to be surveyed and 75.63% of them actually completed the survey. [9]

Appendix Table 3 compares our sample to participants of the 2016 American Communities Survey (ACS), a sample of 3,156,487 individuals from across America. The mean household income for our set of respondents is $47,484, compared with $91,850 nationally. We have a significantly higher fraction of black people – 43.7% in our sample compared to 12.7% in the national average. Conversely, we have a lower fraction of hispanic people – 9.7% in our sample compared to 17.8% in the national average. Our sample is statistically similar to the national average in terms of gender. As one might expect, our sample is less educated than the national average – 48.8% have incomplete college degrees compared to

[9] Appendix Table 4 displays summary statistics on various demographics and poverty level for various samples – (1) phone screener sample, (2) ineligible from phone screener, (3) eligible from phone screener, (4) eligible individuals who agreed to participate and (5) participants who completed the in-person interviews. We also display whether the samples are statistically different from each other in terms of demographics and current poverty status. Compared to participants who did not grow up poor (or, are ineligible for the in-person interviews), participants who grew up poor are more likely to be poor now, are younger, have a higher fraction of black and hispanic people, and are lesser educated. Compared to all participants who were eligible for in-person interviews, participants who finally filled in the paper survey are more likely to be poor now, are younger, more black and less hispanic.

40.7%, while only 20.8% of respondents have a bachelor's degree or higher in our sample compared to 23.1% nationally. In other words, our sample is more likely to be poor, has a higher fraction of black people and is less educated.

We also compare our sample to a sample of individuals from the National Longitudinal Survey of Youth (NLSY) who experienced poverty in their youth. In contrast to the ACS sample, our sample and the subset of those in the NLSY who experienced poverty in their youth look quite similar.[10] Appendix Table 2 compares 4 demographic variables including individual income as well as some questions that assess locus of control and mental illness that are both contained in our sample and the NLSY. Column (1) contains summary statistics from our sample. Column (2) presents these statistics for the sample of the NLSY who experienced poverty in their youth. Column (3) is identical to column (2) but reweights the observations so that the NLSY data has the same distribution on 3 exogenous variables – age, race, and gender – as our sample.

We begin by comparing our sample to the unadjusted NLSY sample. The average age in our sample is 49, and 53 in the NLSY sample. The difference, 4 years, is statistically significant. As NLSY gathers information on individual and *family* income only, we compare individual income across both samples. The mean individual income in our sample is $28,312 compared to $27,751 in the NLSY sample. The difference, $561, is not statistically significant. Other demographic variables such as gender are also statistically similar. Our sample has more blacks and less Hispanics than the NLSY. We also compare questions from the adult mental illness index that overlap between the two surveys. Our sample has worse adult mental illness on three of the four subcategories and lower self-esteem scores than the NLSY.

Comparing our sample to the adjusted NLSY sample gives similar results except on individual income and fraction of black people. Mean individual income in the adjusted NLSY sample is $24,247 and is statistically smaller than the mean in our sample. Mean age in the adjusted NLSY sample is 53 and statistically larger than the age in our sample, while the percentage of women is 52 and statistically similar. While the adjusted NLSY sample has more Hispanics, it has statistically similar percentage of blacks compared to our sample. With regard to adult mental illness and other psychological traits, the adjusted NLSY sample looks similar to the unadjusted NLSY sample i.e. the sample has better adult mental illness on three of the four subcategories and higher self esteem scores compared to our sample.

[10]We say an individual "experienced poverty in their youth" in the NLSY sample if a respondent from 2014 was between the ages of 14 and 22 in 1979 and the survey reported that their family was below the poverty line in 1979.

H. Correlations with Income

To get a sense of how these new data – unprocessed – correlate with income, Appendix Figures 10 - 14 displays binned scatter plots of household income on a set of 18 indices individually, which together, encapsulate all 350 questions asked of our respondents. All figures plot a scatter graph, as well as a quadratic fitted line, of the log of household income on binned categories of indices with each observation weighted by its associated survey weight.

The general picture that emerges is that adult income is strongly correlated with mental illness before 16 years of age, psychological traits, family environment, lifetime traumas and neighborhood mobility.[11]

Appendix Table 5 displays pairwise correlation coefficients between any two given indices and provides a good sense of how potential covariates of intergenerational mobility might relate to each other. Of the several coefficients presented, the most noteworthy ones are between adverse childhood experience and mental illnesses before 16 years of

[11] Appendix Figure 10 displays the relationship between income and four health-related indices: (a) mental illness before 16, (b) physical illness before 16, (c) psychological index, and (d) diet. Mental and physical illness indices were created as totals of standardized responses to questions on "Childhood Health Experience" . Psychological index is the total of 7 sub-indices: resilience, locus of control, growth mindset, grit, self esteem, Big 5 personality traits, and self control. Diet is the total of standardized responses to 2 questions – fraction of days the respondent received three meals and fraction of days he received a balanced diet between the ages of 5 and 12. A detailed description of how each index was created is given in the Online Appendix D.3.

Both mental illness before 16 and the psychological index show strong correlations with income – income declines as mental illnesses in childhood increase and income rises steeply with more psychological skills. Appendix Figure 11 similarly plots six indices which, together, provide a reasonably comprehensive picture of each respondent's childhood family environment. Income is positively correlated with the *quantity* of adult relationships trusted in their childhood, and is also positively correlated with the *quality* of those relationships – individuals who could trust their parents had higher income than those who did not trust their parents but trusted a teacher or coach instead. Family environment – which encompasses true/false statements such as "we fought a lot in our family" , "family members were rarely ordered around" , "we didn't say prayers in our family, "we didn't believe in heaven or hell" , "family members sometimes hit each other" , "everyone had an equal say in family decisions" etc. – is positively related to income and so too, is the quality of a family's social network (defined as the total of standardized responses to questions on if parents' close friends lived in the same neighborhood, graduated from college, worked full time or were a different race). Surprisingly, parenting skills and relationship with parents was not directly correlated with income. These variables are strongly correlated to family environment, however.

Appendix Figure 12 displays the same graphs for five childhood experience indices – (a) adverse childhood experience (ACE), (b) risky attitudes as a teenager, (c) trauma before 18 years of age, (d) any trauma in lifetime, and (e) beliefs about success in life. Income levels decline when teenagers exhibit more risky behaviors or when individuals report more lifetime trauma. Surprisingly, however, the relationship between income and adverse childhood experiences is not significant, though ACE and youth mental health have a correlation of almost 0.5.

The final set of graphs are shown in Appendix e 13 which plot income and three neighborhood indices – (a) neighborhood income mobility, (b) fraction of fathers present in neighborhood, and (c) neighborhood safety index. All neighborhood indices are positively correlated with income levels.

age (0.426), adverse childhood experience and traumas experienced before 18 years of age (0.517) and adverse childhood experience and parenting (-0.572). Good parenting is highly correlated with better relationship with parents (0.510) and, consistent with Chetty et al. (2018), neighborhood mobility is highly correlated with fraction of fathers present in zipcode (0.779).

Traditionally, social scientists interested in correlates of intergenerational mobility have estimated models of the following form:

$$income_{t+1} = \alpha + \beta X + \gamma income_t + \epsilon \tag{1}$$

Western and Pattit (2010) use this approach on the National Longitudinal Survey of Youth 1979 data and find that while two-thirds of non-incarcerated low-income men are upwardly mobile, only one in four out of incarcerated men rises out of the bottom quintile of the earnings distribution. Sanbonmatsu et al. (2012) use data from interviews of adults from the "Moving to Opportunity" households and infer that moving into a low-poverty neighborhood during childhood has substantial effects on the physical and mental health of adults.[12] Most recently, Chetty et al. (2018) demonstrate that the only variables that explain racial differences in mobility across geographies (out of 23 analyzed) is the fraction of black fathers present in the census tract and racial bias in the county (measured as scores on Implicit Association Tests and an index based on the frequency of Google searches for racial epithets). The authors estimate that children who grow up in a tract with 10 percentage points more black fathers present have incomes that are 0.5 percentiles higher on average. Conversely, counties with 1 standard deviation higher level of racial bias against blacks have mean income ranks that are 0.8 percentiles lower.[13]

We also try tried running a "horse race" regression of all our variables, on all three of our income variables. The results are depicted in Figure 1. We plot the significant coefficient from each regression when all variables are normalized, and the standard deviation of the coefficients in absolute value is also normalized to 1, for comparability with out next results. Years of education is the most significant variable in all the specification. Mental health also is significant in all specifications. Other variables vary between specifications, so the results are not particularly robust. Moreover, the sign if often unintuitive, as grit or probability of reaching the top 20% from the 25 percentile in this zipcode enter with a negative sign. This is a known problem is using descriptive methods to assess variable

[12] Appendix E gives a review of studies that estimate correlates of intergenerational mobility from various data sources.

[13] Although, black-white wage gaps are smaller in "good" neighborhoods with low poverty, high rates of father presence and low rates of implicit bias, fewer than 5% of blacks live in such neighborhoods.

importance with data sets which have many potential explanatory factors (Mullainathan, and Speiss, 2017).

Surprisingly however, childhood risky behaviors, physical or sexual abuse or neighborhood mobility do not register as significant once one controls for other variables. Of course, they may be operating through variables such as trust of adults and mental health, but if we were designing an intervention based on these data and were not confident about the income production function, one could conclude that increasing education and giving them resilience training – and ignoring parenting, abuse, or risky behaviors – would significantly increase income.

This inspired us to think beyond the traditional approach.

3 Methods

In what follows, we develop a method that one can use to understand which factors from observational data have the greatest potential to increase a pre-specified outcome. In our particular case, we view these variables as potential levers to change in a field experiment designed to increase income mobility and adult well-being. Our method unearths factors that increase mobility directly, as well as the variables that influence those factors, and so on, without assuming the causal structure among the set of data we collected.

This method does not tell us which specific experiment we should run. It characterizes the joint distribution of observables after the ideal intervention. However, it does not tell us anything about which interventions would get us to this distribution. Nor does it tell us if we should intervene in a variable directly, or intervene in another variable that is causally affecting it. We are left with a reverse engineering problem of designing the intervention that would generate this distribution, that we can only solve using our prior knowledge.

We begin with a simple example we can solve analytically.

3.1 An Example

Let there be only two variables related to chances of escaping poverty: ACE, an indicator for whether a child was abused and $Grad$, an indicator for college graduation. Further, assume a very simple causal structure where $ACE \rightarrow Grad \rightarrow Income$. Specifically, assume that $P(Grad = 1|ACE = 0) = 0.95$ and $P(Grad = 1|ACE = 1) = 0.05$. For income assume that

$$income = \alpha * Grad + \epsilon_0$$

The optimal intervention in this example is to reduce abuse levels, which would increase graduation rates and as a result increase income. This would also be the typical way in which people escape poverty.

Notice however, that conditional on college graduation, income is independent of abuse status. In other words, a child who suffered abuse that graduates from college can expect the same income as a college graduate that was not abused. Traditional methods that don't take into account the graduation production function, will estimate $\widehat{\alpha}$ correctly and no impact for abuse. Namely, estimating

$$income = \widehat{\alpha} * Grad + \widehat{\beta} * ACE + \epsilon,$$

yields $\widehat{\alpha} = \alpha > 0$, $\widehat{\beta} = 0$. This is because these methods are designed to optimize prediction, and given graduation, abuse is unimportant for predicting income. Designing an experiment based on such prediction methods will suggest that we should help kids graduate from college. Childhood abuse doesn't matter, as long as everyone gets to go to college and has the potential to graduate.

However, abused children graduating from college are rare, because of the additional challenges they have to face. In that case, any intervention that lowers the costs for abused children to enroll and ignores the abuse itself (e.g. free test prep or more aggressive guidance counselors) will likely fail because it will not translate into increased graduation. In contrast, if we take into account the graduation production function, we will choose different interventions.

The challenge is that we do not know the causal structure that is generating our data. Therefore, we want to avoid approaches that require knowledge of the real structural model. In the context of this example, we do not necessarily know the link from abuse to college graduation and from graduation from income. Hence, we would not estimate a college production function which depends on abuse level as we have no indication that this is the right model.

To overcome this problem we will use a typically overlooked source of information - the distribution of the X variables. We develop a method that will not just use the prediction of income given graduation and abuse, but also the unconditional joint distribution of abuse and graduation. We will present a set of conditions that, if satisfied, yield the very intuitive results that it is easier to generate things that already appear in the data. Hence, we will look for a combination of X variables that on the one hand predict high income, and on the other will be relatively common in the data. Therefore we will want to increase college graduation rate in order to predict higher income, but also to reduce

abuse levels, since typically and the data they covary.

Notice that while our method will correctly identify that both variables need to change, it will not be able to identify the way to do that. Since we do not know the causal structure, we do not know that the intervention should focus on abuse, which will generate higher graduation rates as a result. We only know that in the optimal intervention both need to change.

3.2 Using Observational Data to Inform Social Experiments

Imagine that we want to improve some outcome Y (say, log income) and we have data on many observables, X, from a set of individuals. We want to design an experiment that would generate the largest expected increase in income. Below, we describe a general method to accomplish this based on the following assumptions.

The first assumption ensures there is a problem worth solving – finding correlates that have no potential causal impact is useless. This assumption is not meant to be a reasonable description of reality, but rather something that will be verified in the field experiment. The following four axioms are the key innovation of our method. It accounts for the causal links between different X variables in a simple manner. We formalize this intuition below.

Assumption 1. *There is a causal relationship from X to Y such that $Y = f(X) + \varepsilon$ and $f(X) \perp \varepsilon$.*

If we assumed f was linear, we could estimate this relationship with ordinary least squares regressions, or, if we want to make fewer assumptions, supervised learning algorithms. These approaches, which are standard in the literature, are problematic for the reasons discussed throughout and will now be analyzed rigorously.

Let there be a process that determines $x \in X$, albeit imperfectly. Mark the distribution of X at status quo by P_0. Define $\mathcal{I}$ to be the set of possible interventions, where an intervention $I \in \mathcal{I}$ changes the distribution of X from P_0 to P_I. We assume $I \perp \varepsilon$. So the causal structure is such that I is affecting X which in turn changes Y

$$I \to X \to Y$$

Let the expected effect of an intervention, I, on an outcome Y, be denoted $\Omega(I)$,

$$\Omega(I) = E[Y|X \sim P_I] - E[Y|X \sim P_0] = E[f(X)|X \sim P_I] - E[f(X)|X \sim P_0]$$

We do not assume that one has any data on interventions so the distribution P_I cannot be directly estimated. Moreover, we do not assume that we have a model of all the potential causal relationships of X, which precludes one from writing down a structural model. And, importantly, we don't assume that we can shape P_I in any way we want.

In lieu of making these typical assumptions, let $C\left(P_I|P_0\right)$ denote the cost of an intervention I. Our goal is to find an intervention I – a distribution of X – that maximizes the expected effect on income, given some budget constraint:

$$\begin{gathered}\max_{I\in\mathcal{I}}\Omega\left(I\right)\\ s.t.\ C\left(P_I|P_0\right)\leq B\end{gathered}$$

Using typical supervised learning methods, or even OLS is equivalent to assume that for every I, $C\left(P_I|P_0\right)\leq B$. When cost is always within budget, the maximization problem is unconstrained. The only goal is to maximize $f\left(X\right)$, so we can estimate $\widehat{f}$ with a supervised learning algorithm and find the combination of X to maximize it. For example, if we believe $f\left(X\right)$ is linear, we can estimate it with OLS, and the coefficients from the estimation will tell us the partial impact of each variable.

But if cost was always within budget, there would have been no science of economics. So instead we replace this implicit assumption with four axioms that we believe are reasonable requirements to have from any cost function.

Axiom 1. ***Continuity:*** *$C\left(P_I|P_0\right)$ is continuous in P_0, P_I (see Appendix A for formal definition).*

Axiom 2. ***Weakly positive:*** *$C\left(P_I|P_0\right)\geq 0$, where $C\left(P_I|P_0\right)=0$ iff $P_I\left(x\right)=P_0\left(0\right)$ almost surely.*

Axiom 3. ***Invariance to Representation:*** *Let P_0, P_I be some distributions on X. Define an invertible function $\Phi: X\rightarrow Z$ that is differentiable almost everywhere. Let Q_0, Q_I be distributions on Z s.t. $P_0\left(X\right)=Q_0\left(\Phi\left(X\right)\right), P_I\left(X\right)=Q_I\left(\Phi\left(X\right)\right)$. Then $C\left(P_I|P_0\right)=C\left(Q_I|Q_0\right)$.*

Axiom 4. ***Additivity:*** *For any two independent interventions I_1, I_2 (see appendix A for formal definition), the overall cost is the sum the individual cost of implementing each of them separately.*

The first two axioms are natural characteristics of any cost function. We require that the cost function is continuous (so similar intervention have similar costs), and that cost is weakly positive.

Axiom 3 is more restrictive. One implication of this requirement is that the cost doesn't depend on representation. Therefore, rescaling the variables or presenting them in a logarithmic scale, doesn't affect cost. This is a necessary property we should expect from every cost function.

The more restrictive implication of this assumption is symmetry between all X variables. It guarantees that the cost does not depend on any particular order or name of the variables.[14] As a result, 1σ increase is equally costly, for all variables that are equally distributed.

An important consequence of this restriction is that it implies our results are not going to depend on any prior knowledge we might have on the cost of some subset of variables. In that aspect, our method is similar to off-the-shelf machine learning methods such as random forests that people typically use to identify which variables are important for prediction, where only the data determines which variables seem more important.

This is a strong requirement, as often we do have some prior knowledge about the relative difficulty of changing each variable. At the extreme case, where one knows the full model that produces the X variables, there is no use for our method, as structural estimation would probably be more suitable. But in many cases we lack the full structural knowledge of cost, but still do not want to assume full symmetry between all our variables. We offer two different ways of doing that which we discuss in Section 5.2.

Axiom 4 pins down the exact functional form of our cost function. While we find this axiom to be a plausible requirement, it is still easy to imagine cost structures that would deviate from it. For instance, there might be some cost that can be shared when implementing two interventions jointly, generating economies of scale. Conversely, it is possible to imagine an increasing marginal cost, that would make the joint implementation more costly than the sum of costs for each one.

In Section 5.2 we discuss different cost functions that only satisfy axioms 1-3, but not 4 and show that they give similar results qualitatively, with slightly different functional forms.

The following Theorem[15] shows that there is a unique function that satisfies all four requirements.

Theorem 1. *Let P_I, P_0 be functions that satisfy regularity conditions. Mark by π_I, π_0 their corresponding densities.*

A function $C\left(P_I|P_0\right)$ satisfies Axioms 1-4 $\iff C\left(P_I|P_0\right) = cKL\left(P_I; P_0\right)$

where $c > 0$ is constant and KL is the Kullback-Liebler Divergence:

$$D_{KL}\left(P_I|P_0\right) = \int \pi_I \log \frac{\pi_I}{\pi_0}$$

[14] Formally, if Φ is a function that only changes the order of the variables, it must maintain the same cost structure.

[15] A similar theorem was proven by Hobson (1969).

Proof: Appendix

This theorem implies that cost is higher for larger changes. The Kullback-Liebler Divergence quantifies the similarities between the two distributions, before and after the interventions. Small changes to the distribution, that keep P_I more similar to P_0 are therefore easier to implement, under these Axioms. Generating rare events (such as sending abused kids to college) creates a larger divergence from the status quo, and are therefore more costly.

With this result we can simplify the maximization problem considerably. First, rewriting it in Lagrangian form

$$L = \Omega(I) - \lambda\left(C\left(P_I|P_0\right) - B\right),$$

where $\lambda = \frac{\partial E[\Omega(I^*)]}{\partial B}$ is the shadow price of an intervention.

We can simplify the Lagrangian further with two observations. First, $E[Y|X \sim P_0]$ is the status quo, which is constant for all I. Hence, to maximize $\Omega(I)$ we only need to maximize $E[f(X)|X \sim P_I]$. Second, B is a constant. Thus, we have the following result:

Proposition 1. *Under Assumptions 1 & A1-A4 , the optimal intervention I, solves the following maximization problem*

$$\max_{P_I} E[f(X)|P_I] - \lambda D_{KL}\left(P_I||P_0\right) \tag{2}$$

for a given value of λ.

Hence, we are looking for small changes that would generate large increases in the outcome variables. When $\lambda \to \infty$ it means we can't change the distribution of X much ($B \to 0$) and P_I would converge to the status quo distribution P_0. When $\lambda \to 0$ it means we can change it easily ($B \to \infty$) and our method would converge to a supervised learning exercise.

The solution to this maximization problem yields the joint-distribution of X after the most cost-effective intervention in expectations. Some of the X variables will be changed in order to increase Y directly. Other variables in X will be changed in order to affect Y through other X variables. And some X variables may also change just as a side-effect of the change in the causal variables. While we cannot distinguish between those different cases without additional knowledge, the solution still focuses our search to a much fewer set of interventions.

While conceptually straightforward, this maximization problem can be difficult to solve in practice. We need to estimate two functions: $f(X)$ and P_0. Estimating $f(X)$ is a standard problem for which one can apply an array of statistical methods ("supervised

learning"). To estimate P_0, we need to characterize the distribution of the Xs without any outcome variables; an "unsupervised learning" problem. Finally, we need to solve the maximization problem and find P_I for those estimated values.

In the next sections, we solve the maximization problem under three different sets of assumptions. First, we estimate P_0 assuming X has a multivariate normal distribution and estimate f assuming it is linear. The solution in this case is quite simple. We then use empirical likelihood to estimate both P_0 and f non-parametrically which provides a solution to our maximization problem allowing for flexibility in the relationship between the variables and the relationship between those variables and our outcomes. Finally, we allow for heterogeneity in the distribution of X across individuals.

A. X IS NORMAL, f LINEAR

Assume X has a multivariate normal distribution and f is linear. With these assumptions, we can prove the following result – which greatly simplifies our problem.

Proposition 2. *Assume $f(X)$ is linear, $P_0 = N(\mu_0, \Sigma_0)$ and Assumptions 1, A1-A4 hold. Assume also that X is normal after the intervention ($P_I = N(\mu_1, \Sigma_1)$). Then for the optimal choice of I*

$$\mu_1 = \mu_0 + \rho$$

$$\Sigma_1 = \Sigma_0$$

where

$$\rho = \frac{1}{\lambda} COV(X, Y)$$

This proposition simplifies the characterization of the optimal choice for I. First, if X is normal and $f(X)$ is linear, the optimal intervention is an increase of X by a constant vector ρ. Second, this constant is the covariance of X with Y divided by lambda. If we standardize X and Y (to make them unit free) then it's proportional to the correlation coefficient. Thus, raw correlations provide the direction, and if the shadow price is smaller, we proceed further in that direction. And, given our focus on the relative importance of each variable compared to other variables, the value of λ doesn't effect the results. This is a unique case, where we can analytically solve for the optimal intervention without actually estimating f and P_0 directly.

The intuition described in our earlier example shines through: this method may choose to increase some variables even if they do not effect the outcome directly, once controlling for other variables. In our earlier example, ACE would have not been picked up by OLS once controlling for $Grad$, but still would have been picked up by raw correlations.

B. Non-Parametric Estimation of P_0, f

The assumptions that X has a normal distribution and f is linear are quite strong. We now outline an approach to estimating both P_0 and $f(X)$ non-parametrically, using empirical likelihood (Owen, 2001).

To estimate P_0 we assume that every x_i we observe in the data has a probability of $\frac{1}{N}$, where N is the sample size. This is the distribution one assumes on the data when using bootstrap methods. The probability to observe an x that is not in the data is set to 0. In symbols:

$$\pi_0(x) = P(X = x | I = 0) = \begin{cases} \frac{1}{N} & \exists x_i \in data \text{ s.t. } x_i = x \\ 0 & \nexists x_i \in data \text{ s.t. } x_i = x \end{cases}$$

Thus, an intervention alters the probabilities of the observed X vectors. The KL divergence for any choice of I is $D_{KL}(P_I|P_0) = \sum_x \pi_I(x) \log \frac{\pi_I(x)}{\pi_0(x)}$.

For f we estimate $\widehat{f}$ as

$$\widehat{f}(x_i) = \log w_i$$

where x_i is the value of X for the ith observation and w_i is income for that observation.[16]

Our problem simplifies to choosing values for $p_i = \pi_I(x_i)$ where x_i is the ith observation. The solution is attained from the following proposition:

Proposition 3. *Assume P_0, f are distributed as above and Assumptions 1, A1-A4 hold. Then for the optimal choice of I*

$$\pi_I(x_i) \propto w_i^{\lambda^{-1}}$$

Therefore, the solution is a reweighted distribution, that puts larger weights on higher-income individuals. Notice: higher λ means that we can't change much and probabilities remain similar to uniform. Lowering λ puts more weight on high-earners. We use a value of $\lambda = 100$, though different values yield similar results.[17]

C. Heterogeneity

What if it's not actually possible to make all people to look more similar to the typical high earning people? In this case, it's more sensible to try to make people to be more similar to people that have higher income but are more similar to them on other dimensions.

[16] If $\pi_0(x) = 0$ (which means that x is not observed in the data), then $\pi_I(x) = 0$, otherwise $D_{KL} = \infty$. Therefore we only need to define $\widehat{f}$ on observed values.

[17] We also found that higher values of λ yield results more similar to ones we get under linearity assumptions.

To put this in the context of our framework, we assume that there is heterogeneity in the distribution of X. Hence, it's possible that high earning people are drawing X from a different distribution. This would mean that drawing such Xs could be much more costly for some people. Formally, we will assume that $P_{0,i}$ is different for every observation i. Moreover we will assume that this probability is higher for neighboring values: other values of x we observe in the data for observations that are close. As a result, the cost $C_i(I) = C\left(P_{I,i}|P_{0,i}\right)$ is lower when we try to change x_i to its neighbors value.

Assumption 2. *Every individual i draws X from the following distribution:*

$$P_{0,i}\left(x\right) \propto \begin{cases} \exp -\frac{dist(x,x_{i'})^2}{2\sigma^2} & \exists x_{i'} \in data \text{ s.t. } x_{i'} = x \\ 0 & \nexists x_{i'} \in data \text{ s.t. } x_{i'} = x \end{cases}$$

and $dist\left(x, x_{i'}\right)$ is Mahalanobis.

The parameter σ will set how much we penalize for distance. When $\sigma \to \infty$ we can turn child i to any other child in the data with equal costs, and so there's no heterogeneity. As $\sigma \to 0$, we can only change to the closest neighbor. There is a tradeoff between bias and variance in the choice of σ. High σ will use all data, and would therefore be more biased but with less variance. Low σ will use fewer and closer data, and will therefore be less biased but noisier. We use $\sigma = 1$, as high values of σ generate results that are similar to our non-parametric methods. For the choice of f we will use the same non parametric method we used before and set $\widehat{f}\left(x_i\right) = \log w_i$.

We can limit ourselves to intervention that sets some positive probability $\pi_{i,j} = P_{I,i}\left(x_j\right)$ where x_j is the jth observation in the data, and $\pi_{i,*}$ is the specific probability distribution of the Xs for child i after the intervention.[18] Our goal is then to choose values for π_{ij} for every j s.t. $\sum_j \pi_{ij} = 1$.

Proposition 4. *Assume P_0, f are distributed as above and Assumptions 1 & A1-A4 hold. Then for the optimal choice of I*

$$\pi_{ij} \propto w_j^{\lambda^{-1}} \exp \frac{dist\left(x_i, x_j\right)^2}{2\sigma^2}$$

Intuitively, this exercise is similar to the reweighting in the non-parametric section. The key difference is that now we put more weight on closer neighbors. Therefore, high-income people who look very different from the rest of our data would get a lower weight, compared to the non-parametric case. This captures the intuition that people that look

[18] Any positive probability on an unobserved value for x would yield an infinite cost.

very different, might have their Xs produced in a different way, and therefore an intervention that would try to make all people more similar to them would be less likely to be successful. Overall, under the optimal intervention we would have a distribution of X with the same support of our data, that has a higher probability to draw X values of high income people, who are similar to the X distribution in the data.

We choose $\lambda = 100$ as we chose in previous sections. We calculate π_{ij} up to a constant, using the above equation, and normalize to get the probabilities sum to one. This gives us a distribution to draw each value of x_j which is $P_I(x_j) = \frac{1}{N}\sum \pi_{ij}$.

D. Adding Weights and Controls

Our method can accommodate control variables. Assume some moment of the distribution $m(X)$ cannot be changed in any intervention. In this case, we restrict the maximization problem (2) only to P_I that satisfy $E[m(X)|P_I] = E[m(x)|P_0]$. i.e, distributions of X where that moment does not change. We can also accommodate sample weights, such that the status quo distribution $\pi_0(x)$ is not uniform. This is summarized in the following proposition.

Proposition 5. *Assume some $\pi_0(x)$ with finite support, $f(x_i) = \log w_i$ and Assumptions 1 A1-A4 hold. Assume also that a vector of moments of X marked by $m(X)$ is fixed such that $E[m(X)|P_0] = E[m(x)|P_I]$ then*

$$\pi_I(x_i) \propto \pi_0(x_i) w_i^{\lambda^{-1}} \prod_j \exp\left(\frac{\gamma_j}{\lambda} m_j(x_i)\right)$$

Note that π_0 could be defined using sample weights. γ_j values would be chosen such that $E[m_j(X)|P_0] = E[m_j(x)|P_I]$. In practice we will define $m(X)$ to include indicators for each race and gender, age and parental income index. A similar proposition also exist for the case of heterogeneity, and we will prove the more general case of this proposition in the appendix.

3.2.1 Statistical Package

We are not the first to think about the problem of designing an experiment from wide data. Moreover, we hope that more researchers would adopt this approach in more contexts when they try to decide between a large number of interventions that could, potentially, increase some important outcome variable.

We developed an R package that implements all the aforementioned methods at great ease. The package finds P_I for a choice of λ and one of the three methods above. It

calculates standard errors and significance levels and produces the same plots as in this paper.[19]

4 Results

In this section, we present results gleaned from implementing the methods described above on our new set of data. Each one of the methods yields a new distribution of P_I, which we compare to the original distribution of the data P_0. Our primary goal is to detect which variables undergo the biggest changes. We do this by plotting a series of figures with the change for each variable in rank order, for all variables in our dataset that are statistically different from zero.

To account for non-linear, or even non-monotonic changes, we compare the full CDF of each variable under both the distribution in the data P_0, and the calculated distribution for the optimal intervention P_I. Let $q_p\left(X_j|P_z\right)$ be the p-percentile of the variable X_j , under distribution P_z (where $z \in \{0, I\}$). We calculate

$$\tau\left(X_j\right) = \frac{1}{100}\sum_{p=0}^{100}\left(q_p\left(X_j|P_I\right) - q_p\left(X_j|P_0\right)\right)^2$$

Note that under the assumption of linearity and normal distribution (Proposition 2), $\tau\left(X_j\right)$ is simply the difference in means. We standardize each X_j variable such that $V\left(X_j|P_0\right) = 1$ to make this unit free.

For comparability across methods, we standardize the distribution of $\tau\left(X_j\right)$ i.e. we divide $\tau\left(X_j\right)$ by its standard deviation across all variables in a given method:

$$\frac{\tau\left(X_j\right)}{\sqrt{\sum_{j'}\left(\tau\left(X_{j'}\right) - \overline{\tau\left(X_{j'}\right)}\right)^2}}$$

One can interpret the x-axis in each figure as the units of standard deviation for $\tau\left(X_j\right)$. T We calculate p-values using a permutation test and display only variables that are significant at the 10% level.[20] The bars surrounding each coefficient estimate is the 90% confidence interval, calculated using bootstrap.

[19]The package is available to download through CRAN at https://cran.r-project.org/web/packages/optinterv/index.html

[20]Specifically, we draw 1000 permutations of the outcome variable, and calculate $\tau\left(X_j\right)$ for each permutation. Under H_0, where the distribution of X_j doesn't change ($P_I\left(X_j\right) = P_0\left(X_j\right)$), the rank of $\tau\left(X_j\right)$ in the real data should be uniformly distributed within its value under permutations. Hence the p-value is simply the share of permutations that generate a higher value of $\tau\left(X_j\right)$.

4.1 Standard Methods

Before displaying the results from our methods, we start from describing the results from commonly used descriptive methods. At the end of Section 2 we discussed our results from an ordinary least squares (Figure 1). Overall, the results of this exercise seem unreasonable. Seemingly important variables (e.g. abuse) are missing, the signs of some crucial variables are unexpected and the results are not robust across different definitions of income. While these variables are the most important for predictions, it is unlikely that they are the most important variables when designing an intervention.

A perhaps more common way to identify important variables is using Random Forests. Specifically, after training a random forest model $\widehat{f}_{RF}$, one can see which variables were most commonly used in the constructed trees, or which variables have the largest effect on the predicted outcome. While this practice is recommended in the context of finding variables important for prediction, it is less clear that it is useful for designing experiments.

The results from this exercise are plotted in Figure 2. We plot the top 10 most important variables, using different definitions of income. The first thing to notice is that the results are quite unstable across the different definitions. This may allude that this method is not robust. Years of schooling, for instance, is the most important variable for HH and adjusted HH income but not in the top 10 of the individual income. As with the OLS, Adverse childhood experience does not seem to be particularly important in any of the income definitions.

Contrastingly, we find two neighborhood variables to be among the top 10 variables in each of our setting - probability of reaching the top two deciles given parents in percentile 25, and fraction of father present. Those variables obviously have high predictive power, even when other variables are given. However, once again, this does not mean that these variables are also the most promising when thinking about interventions. Put differently, neighborhoods with high chances of escaping poverty are quite rare. Therefore, it might be more promising to focus on the reasons why within neighborhoods, so children are able to escape poverty, while others don't.

4.2 Income Results

We begin with log household income as our outcome variable and our preferred specification – non-parametric estimation of P_0 and f. Panel A in Figure 3 shows corresponding results. Using non-parametric estimation methods, the most important correlate of income mobility is education. This is consistent with a large literature on the importance of

the quantity of education on income (Card, 1999; Garces et al., 2002; Belfield et al., 2006; Barnett and Masse, 2007; Heckman et al., 2010; Heckman et al., 2013; Elango et al., 2015; Heckman et al., 2016). A close second – and statistically indistinguishable – is resilience. Recall, resilience is the ability to bounce back from stressful situations and is measured by responses to questions such as "It does not take me long to recover from a stressful event".

Surprisingly, half of the significant correlates of intergenerational income mobility are psychological skills: resilience, Big 5, self-esteem, self control, locus of control and grit. Relatedly, mental health problems is also significant in this specification. Other important variables are whether the respondent was ever in trouble with the police in their youth, had adverse childhood experiences, and the existence of adult relationships they trusted.

If we adjust household income by household size (Figure 3b), or use individual income we get similar results.[21] For individual income we get that risky attitudes as a teenager, whether the respondent lived with a mother when they were young, and growth mindset – another psychological skill are also significant.

These results are in contrast to much of the literature on the correlates of income mobility, though consistent with Nyhus and Pons (2005), Heckman et al. (2006), Currie and Widom (2010), Mofitt et al. (2011), and Heckman et al. (2013). For instance, we do not find that church going, fraction fathers present in a zipcode, or mobility indices more generally are significant correlates. Generally, there is a larger focus in our results on psychological skills and the ecosystem embodied by children when they are young, which includes interactions with police and other adverse childhood experiences, risky behaviors, and the adults in a child's life

The results when using parametric assumptions are similar. In Figure 4 we show the results when assuming a linear f and normal P_I, which is equivalent to Pearson correlation. Because of the parametric assumptions we have more statistical power which leads to more significant variables. Again, under these assumptions, we get all the variables that we get above along with family environment, fraction of fathers present in a zipcode, neighborhood safety, neighborhood mobility, parenting and family network. Similar to above, of the top eleven variables, seven of them are specific psychological skills. The Spearman (rank) correlation between the non-parametric method of estimating P_0 and f and assuming that X is multivariate normal and f is linear is .86 (Table 1).

Figure 5 shows results when we account for heterogeneity. All but two variables –

[21] We adjust income for household size using the same method employed to calculate the Census Supplement of Poverty Measure (U.S. Census Bureau, 2017). A detailed description of the adjustment is provided in the Data Appendix.

self-control and locus of control – from the non-parametric approach continues to be significant. Additional correlates that become significant are family environment and diet. Note that The correlation between the non-parametric approach and an approach where we account for heterogeneity is 0.82.

Different methods offer modest differences in the specific variables that are gleaned to be significant. But the general pattern is robust. Table 1 displays the Spearman rank correlations between the various specifications. The correlation between our non-parametric method and random forest is .64 and only .45 with OLS showing larger differences between the methods. Education, psychological skills, trouble with the police, and adverse childhood experiences are *always* significant, independent of assumptions on f and P_0. This is consistent with recent evidence that education centric interventions designed to increase income among the poor – so-called No Excuses charter schools – may have little impact on mobility (Dobbie and Fryer, 2016). And, suggests that these interventions would be more successful if one simultaneously worked to increase psychological skills and the number of adults trusted, and reduce trouble with the police and other adverse childhood experiences.

Analysis of Subsamples

Appendix Figures 15-16 explore the sensitivity of our income correlates across a variety of subsamples of the data. We report our estimates that calculate the mean value of X after the optimal intervention, separately for each group.

For each division we plot the variables that are significant for both groups, and the variables that are significantly different between groups. Since moving to subsamples reduces our statistical power, we use 80% confidence intervals.

For race, we find that four variables are significantly different between blacks and whites. We find that whether one could trust any adults in their childhood is operating in opposite directions for blacks and whites; while it is increasing black's income, it is decreasing in white's income. Similarly, number of adults trusted has a strong correlation with blacks incomes. Although we are not certain about the mechanism that drives this difference, we see that for all respondents that said "yes" to trusting adults in their childhood, whites had a significantly higher fraction of unemployed fathers than blacks. So white respondents may have depended on fathers that were untrustworthy which could have resulted in lower adult incomes.

With subsamples based on gender, we find that family network seems to be much more important for girls. Repeating a grade seems more detrimental for boys. Other variables have statistically similar impacts on boys and girls.

4.3 Adult Well-Being

Thus far, we have concentrated on variants of household and individual income as outcomes. In this section, we explore a wider definition of adult well-being by including adult physical health, mental illness, marital status and alcohol and drug abuse. We will also look at college graduation as an outcome. Results are in Figures 6 and 7.

Mental Illness

The correlates of mental illness are quite different from those correlated with income. Unsurprisingly, six of the top eight correlates of adult mental health are psychological skills measured in adulthood. It is unclear what this means. Ideally, the psychological skills would be similar in childhood and thus interventions on those variables may prevent adult mental illness. It is also plausible however, that whatever is associated with adult mental illness is also associated with lower values of general psychological skills.

More interesting, risky attitudes as a teenager and mental health as a teenager are also associated with adult mental health. Education, physical illness in childhood and adverse childhood experience are also associated with adult mental health in the expected directions. Fathers present in zipcode, whether the respondent lived with their father as a child, trouble with police, family environment, and neighborhood safety are also all significant correlates of adult mental health. The similarity between these results and those from the Moving to Opportunity experiment are striking. In that experiment, moving poor individuals to less poor neighborhoods was associated with significant improvements in mental health.

Alcohol and Drug Abuse

The correlates of alcohol and drug abuse are interesting and intuitive. The most important correlates are risky attitudes as a teenager and self control, followed by trouble with police, mental health before sixteen years of age, and grit. Other variables include various parenting variables and family environment and neighborhood safety. It is important to note, education is not in the top 10 of most important variables. The variables above are more important if one wants to reduce drug and alcohol use.

Physical Health

The top correlates of adult physical health are surprisingly intuitive. The variable physical illness before 16 is now among the top variables, as there is some correlation between physical illness in childhood and adulthood. Good diet is also associated with better adult health. Just as with adult mental illness, adult physical health are also highly corre-

lated with psychological skills, and childhood mental illnesses. Other top correlates are education and risky attitudes as a teenager.

College

Our results for college graduation are similar to our income results with some expected differences. For this exercise we do not use the years of schooling variable since it captures our outcome variable. We replace it with high school completion, which remains the most important variable. Psychological skills, and trouble with the police are still dominant in this exercise, similar to our income exercise. Using college graduation as our outcome captures more variables related to trust in adults and neighborhood quality.

Marital Status

The top correlates for marital status our psychological skills. Seven out of our top nine correlates including all the top five are psychological skills. Another psychological skill that appears only in this specification is patience. Neighborhood variables are more dominant in this specification, and mostly fraction of father present.

5 Testing the Method

Our analysis is, at most, speculative. Whether the bundle of variables we view as significant correlates can increase income is an experimental question. Will it work? We can't know in the abstract. In this section we will try to test our method in three different ways. First we will use historical wide data on charter school performance and examine whether the optimal intervention we predict preforms well, based on experiments that have been run. Second, we test what happens to our results under violation of our axioms. Finally, we preform a cross-method comparison on specific examples that we can solve either analytically or using simulations.

5.1 Validation with Historical Data from Charter Schools

With the benefit of hindsight, we can use our method on the detailed within-school data collected in Dobbie and Fryer (2013). We can then compare the suggested optimal intervention from our method to the intervention implemented in fryer (2014).

Dobbie and Fryer (2013) study data collected from 39 charter schools and correlate it with estimates of school effectiveness. They find that traditionally collected input measures – class size, per pupil expenditure, fraction of teachers with no certification, and the fraction of teachers with advanced degree – are not correlated with school effectiveness.

In contrast, policies suggested by qualitative research – frequent teacher feedback, the use of data to guide instruction, high-dosage tutoring, increased instructional time, and high expectations – explains 45% of the variation in school effectiveness.

Using the same data, we implement our method. The results are interesting and are shown in Appendix Figure 18. The significant correlates of high quality schools are teacher feedback, instructional time, high expectations and high quality tutoring. Interestingly, non-certified teachers are strongly negatively correlated with school effectiveness while data driven instruction is not significantly correlated with school effectiveness.

Fryer (2014) implemented 4 of these and demonstrated large impacts in math and less in reading, a pattern closely resembled in the achievement-increasing charter schools they were gleaned from. Whether Fryer (2014) would have had significantly different results if they had removed non-certified teachers as the suggested correlates from our method shows, is unknown. However, it is important to note that given that literature suggests that certified teachers, better teacher feedback, higher instructional time, high expectations and high quality tutoring have positive partial derivatives with respect to achievement and the data suggests that they are positively correlated with each other, we may reasonably expect an intervention using the suggested correlates from our method to have a positive effect on math and reading scores.

5.2 Violation of the Axioms

In this section we will discuss what will happen if our axioms will be replaced with other, weaker axioms. The first two axioms (continuity and weakly positive) are standard for a cost function. The third axiom, Invariance to Representation (IR), would be violated, if some variables have larger costs than others. We develop a robustness test to check how results would change in this case. We use this method to show that our psychology variables remain important even if they are 20 times more expensive than what we originally assumed. Finally, we test the robustness of the additivity assumption, and show that removing it would generate similar solutions, with different functional form.

5.2.1 Excessive Costs

The axiom of irrelevance of representation assumed that the cost of changes, depends only on the difference between the distributions. While this assumption may hold in expectation, without any prior information on the data, in practice some changes to the distribution are going to be harder than others. For instance, a change even small, in the distribution of height is probably much more costly than a change in basic parental

practices such as bath time. Hence, it is important to verify that our results still hold even if the changes we encourage are more expensive than the method suggests.

Formally, assume that the optimal intervention recommends changing some moment of the distribution. We mark this moment by $m(X)$, which could be any expectation of the X variables such as average year of schooling or an index of the psychology variables, etc. Our concern is that if the cost of changing the moment $m(X)$ is higher than our method predicts, our results would change.

We can test this by solving a modified maximization problem. Without loss of generality, assume that $E[m(X)|P_0] = 0$. The following equation assumes that there is an excessive cost of modifying $m(X)$

$$\max_{P_{I_\alpha}} E[Y|P_I] - \lambda D_{KL}(P_I|P_0) - \alpha |E[m(x)|P_I]| \tag{3}$$

This maximization problem is identical to our original maximization problem at Equation 2, adding an additional regularization component for the value of $m(x)$ after the intervention. This regularization generates excessive cost for any change in $m(X)$. The level of α sets that cost relative to the standard cost that we estimate with the KL-divergence. When α is zero, the IR axiom holds. When $\alpha \to \infty$, this moment is impossible to change.

The solution to this problem is very similar to our original solution of Equation 2. Under the same distributional assumptions of our non-parametric method, the solution is

$$\log \pi_I(x_i) = \log \pi_0(x_i) + \lambda^{-1} \log w - \lambda^{-1} \alpha m(x_i) * sign(E[m(x_i)|P_I])$$

This generates a similar reweighting procedure. $\pi_I(x_i)$ is as it was without regularization, only multiplied by $\exp\left(\pm \frac{\alpha}{\lambda} m(x_i)\right)$. Intuitively, if $E[m(X)|P_I]$ is positive (negative) this would reduce the weights from observations with high (low) value of $m(X)$, which would decrease the expected value $E[m(X)|P_I]$.

Since we do not know the right value for α, we will examine how our results change for various levels of excessive cost. We look at the excessive cost relative to the basic cost, which are both functions of α. Let m_α be the expected value of $m(X)$ in the solution to Equation 3 ($E[m(X)|P_{I_\alpha}]$). The additional cost of changing $m(X)$ is simply $\alpha |m_\alpha|$. We define the basic cost as the minimum cost to change moment $m(X)$ to m_α. Formally this is

$$BC(m_\alpha) = \min_{P_m} \lambda D_{KL}(P_m|P_0)$$
$$s.t.\ E[m(X)|P_m] = m_\alpha$$

Note that this cost is for the overall change in the distribution, including in other variables that are changing. The excessive cost for α is then defined as the ratio between the

additional and the basic cost:

$$EC(\alpha) = \frac{AC(m_\alpha)}{BC(m_\alpha)} = \frac{\alpha |m_\alpha|}{\lambda D_{KL}\left(P_{m^*_\alpha} | P_0\right)}$$

A value of n implies that the additional cost from changing $m(X)$ equals n times the basic cost of changing it to the same level, under our original axioms, and so this moment is $n+1$ more expensive than our original assumptions.

Excessive cost of one moment could affect other moments as well. If moment $m(X)$ is hard to change, the optimal policy could try to change other moments that are close substitutes. For instance, if psychological variables are harder to change, the optimal policy might focus more on neighborhoods variables or parental practices. Other moments could also be complements, and so would change less. It is possible that if psychological variables are harder to change, years of schooling will be reduced as well since it is harder to improve schooling level without changing psychological characteristics.

We use this method to test the robustness of our key finding, that interventions should focus on psychological characteristics. We define $m(X)$ to be a simple average of the psychological variables in our data: Self control, resilience, locus of control, growth mindset, grit self esteem and the IPIP index. We plot in Figure 8 the excessive cost $EC(\alpha)$ against the values of the psychology index, years of schooling and neighborhoods (probability of getting to top 20 percentile from bottom 25) for that α. We pin down the value of the effect on the psychology index to 1 for $\alpha = 0$, and plot the changes in other moments relative to that.

While the impact on the psychology index would decline with excessive cost, it is still substantially high. The results at 0% are identical to our results from our main specification. In this case the impact on psychological index is four times larger compared to years of schooling, since it incorporates all relevant psychological variables. As expected, when the excessive cost of changing the psychological index grows, the impact of the optimal intervention on it declines and approaches zero. Yet even when the excessive cost is eight times more expensive that the regular cost, the impact on the psychology index is still larger than the impact on years of schooling. The importance of neighborhoods slightly increases, but the sign is negative and the the change is not significantly different from zero.

This exercise is useful to show the robustness of the results under varying options for excessive costs. In some cases, however, we might actually have a good understanding of the cost of some subset of variables which we could incorporate. In Appendix B.1 we show that under some regularity conditions, we could incorporate our prior knowledge

on the cost of a subset of variables, and use our method to asses the cost for the rest of the variables.

5.2.2 Violation of Additivity

The additivity axiom(A4) pins down the specific functional form. Excluding it, will yield a general f-divergence (Csiszer, 1963). That is, a cost function between two distributions that can be written as

$$D_f\left(P_I|P_0\right)=\int f\left(\frac{\pi_I\left(x\right)}{\pi_0\left(x\right)}\right)\pi_0\left(x\right)dx$$

where $f\left(1\right)=0$ and $f''<0$. The additivity assumption sets $f\left(u\right)=u\log u$, yielding the KL divergence (up to a constant). We prove that under such a cost structure, the optimal intervention would always generate a reweighting of the status quo distribution, where higher weights would be put on individuals with higher outcomes. Formally,

Theorem: If axioms 1-3 hold then P_{I^*}, the distribution of X after the optimal intervention is a reweighting of P_0 such that

$$\frac{\pi_I\left(x_i\right)}{\pi_0\left(x_i\right)}>\frac{\pi_I\left(x_j\right)}{\pi_0\left(x_j\right)}\iff y_i>y_j$$

Hence, the three axioms, continuity, weak positivity and irrelevance of representation are sufficient to a guarantee a set of solutions that is qualitatively similar, in which probability increases for high incomes and decrease for lower incomes.

The exact reweighting function depends on additional functional form assumptions. While additivity generates a nice elegant solution, other functional forms are possible. In Appendix B we review the solution under alternative choices of f-divergence (different choices of f).

5.3 Cross-Method Comparison

In Section 3 we have shown conditions under which our method will estimate the distribution of X after the optimal intervention. However, this result doesn't guarantee that using our method to design an experiment will in practice provide better results for two main reasons. First, the axioms we outlines may not hold. Second, while optimally we would like to know the list of variables in which we want to intervene, our method only provides a list of variables that will be affected in an intervention. Hence, our method is likely to capture "side effect" variables, that do not have any causal impact on the outcome variables.

In this section we will analyze the conditions in which our methods outperforms common descriptive methods using both simulations and theoretical analysis. We find that interventions that are based on our method are typically more effective, and the gap is larger when there are more indirect causal factors and less side-effects. We also find that our non-parametric method performs substantially better than raw correlations, in non-linear settings.

We simulate four types of X variables. First, variables that have a direct causal effect on Y, which we mark by X_D. Second, variables that are only affecting Y indirectly, through their effect on X_D. We mark those variables with X_I, such that for a vector of random variables δ

$$X_D = \varphi(X_I) + \delta$$

Our third category of variables are side-effects, X_S. These variables are causally affected by X_D, but have no impact on Y. Finally, we have noise variables X_N that are orthogonal to all other variables.

Notice that in this setting, only the variables that directly affect Y are valuable for prediction. Formally, $E[Y|X] = E[Y|X_D]$. Hence, methods that rank variables based on their contribution to prediction will only highlight variables included in X_D. If our axioms hold, an intervention that targets only X_D variables will preform poorly when δ is small.

Theorem 2. *Define I_D as the optimal intervention that targets only X_D within the budget constraint, holding the distribution of X_{-D} fixed. Under Assumptions 1, Axioms 1-4 and regularity conditions (see Appendix)*

$$\lim_{\delta \xrightarrow{p} 0} |\pi_{I_P}(x) - \pi_0(x)| = 0$$

and

$$\lim_{\delta \xrightarrow{p} 0} \Omega(Y) = 0$$

Even though X_D has a causal impact on Y, when δ is small, the intervention would fail to change the distribution of X_D without targeting the indirect variables X_I. This is a more general case for the example we presented in section 3.1, where an attempt to improve college graduation fails without an intervention that deals with abuse levels.

To test what happens when our Axioms do not hold, we use simulations. We start with a case without side effects. We simulate a data generating process with 30 X variables, from which 10 have a direct impact, 10 with indirect impact and 10 are noise. We simulate 100 different non-linear processes, each with 1,000 observations, similar to our actual data (see Appendix C for exact details). We compare four methods for ranking vari-

ables: absolute value of Pearson correlation (our normal-linear case), our non-parametric method, largest coefficients in OLS and variable importance of Random Forest. For each ranking, we test the impact of an intervention that is based only on the top n variables for $n = 1, ..., 30$. Since we simulated the data ourselves we can also find the optimal intervention that uses only n variables for each n.

Figure 9a plots the results. We divide the impact of each intervention by the impact of the optimal intervention, such that all results are on a scale of 0-1 where 1 is the optimal intervention. All interventions are the same once all 30 variables are used, hence all methods eventually reach 1. But our goal is to find a subset of variables to target, hence we will compare how these methods preform under smaller n.

Our non-parametric method outperforms the other methods for most choices n, and especially when $n \geq 10$. The linear methods (OLS and Pearson correlations) preform poorly in this non-linear setting. An intervention that is focused on the top 10 variables from these methods would generate almost no improvement in the outcome variable. Random Forests generates more similar results to our non-parametric method, but mostly for lower n.[22]

We next test what happens when we increase the number of side effects. Figure 9b plots the the performance of our method compared to random forest. Both methods become less precise once we increase the amount of side-effect variables. Our method is significantly more useful when the number of side effects is small and when the number is large, the two methods preform similarly.

The final simulation exercise demonstrates that our method is more useful in cases when it is harder to change the X variables. When the X variables, or at least the variables that enter the production function f are easily malleable, cost is essentially unconstrained, and supervised learning methods that estimate the function f suffice. To show this, we divide our direct impact variables X_D into two groups X_{D_1}, X_{D_2}. We assume X_{D_1} are still set by the indirect variables X_I as before, but X_{D_2} can be directly modified. We simulate the final outcome Y as a weighted average of two non-linear functions for both sets of variables

$$Y = \log\left[(1-\alpha) f_1\left(X_{D_1}\right) + \alpha f_2\left(X_{D_2}\right)\right] + \varepsilon$$

[22]This result is consistent with our previous findings. The relatively good performance of variable importance with random forests is driven by its randomness. Variable imporiance measures the number of times rach variable was used. Since the algorithm selects a random subset of varibles for each tree, the variables chosen are not necssarily the most important variables for predictions. X_I variables that are only operating indirecly, would still be used in random forest since they are associated with the outcome variables. Indeed we find that random forests preforms better, when we increase its randomness such that it is calibrated to choose smaller subsets of variables.

Figure 9c plots the results for various values of the weight α. As expected we find that our non-linear method preforms better than random forest when α is small. In these cases, variables with indirect effect matter more, and our method since our method ranks them higher, it performs better. When α is larger, the outcome Y is set mainly by X_{D_2}, which can be changed directly. Unsurprisingly, for these values, random forest preforms better than our non-parametric method. Yet, our method still preforms well and generate outcome that are close to optimal in this case as well.

Overall, this exercise demonstrates the intuition we have outlined in this paper. Using prediction methods to design experiments implicitly assume that variables can be changed in any manner. Yet, when variables are set in a complex way with more interdependencies, our method outperforms prediction methods. This exercise also demonstrate the advantage of the non-parametric method over methods that assume linearity. While in linear cases these methods would yield similar results, in non-linear cases the non-parametric method strongly dominates.

6 Discussion

Our analysis of intergenerational mobility has developed a new set of facts. We collected new data on individuals who were reared in poverty. This data gives a wide picture on the experience of childhood poverty as it covers many different aspects of life for the same group of people.

We then develop a new descriptive statistical method to assist the design of future experiments, whose goal is to improve some outcome variable. As most descriptive methods are designed for the purpose of predictions, they ignore the potential difficulty of changing the covariates in an interventions. Our method, uses information from the joint distribution of the covariates, and instead of recommending experiments that would extrapolate into cases rarely see in the data, it focuses on the way in which the outcome variable (in our case, escaping poverty) is higher in real life.

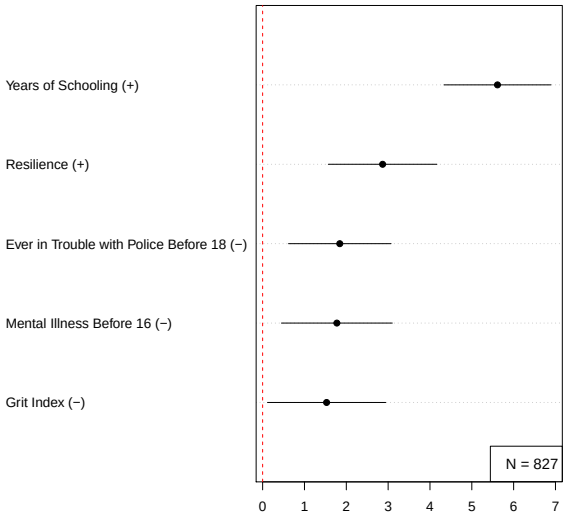
Using our newly collected data and new method, we argue that to increase income among the poor we need a multi-pronged strategy that focuses heavily on the ecosystem children inhabit, their psychological skills to navigate the situations they endure, and, importantly, education. These results are suggestive. We caution against a rush to policy, but rather a rush to experimentation with the goal of boosting income among those who are born poor.

Tables

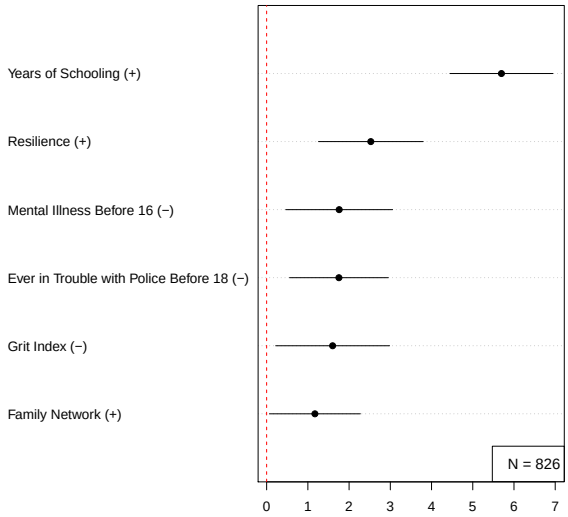
	Non-Parametric	Correlation	Nearest Neighbor	OLS	Random Forest
Non-Parametric	1				
Correlation	.86	1			
Nearest Neighbor	.82	.64	1		
OLS	.45	.38	.49	1	
Random Forest	.64	.75	.57	.28	1

Table 1: Correlation Between Methods - HH Income

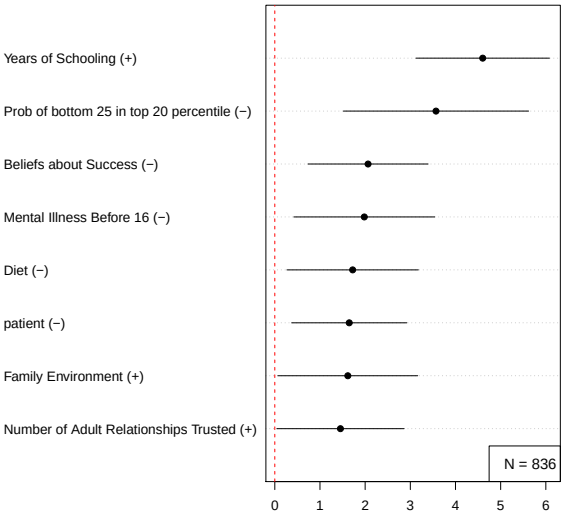
Figures



(a) Household Income

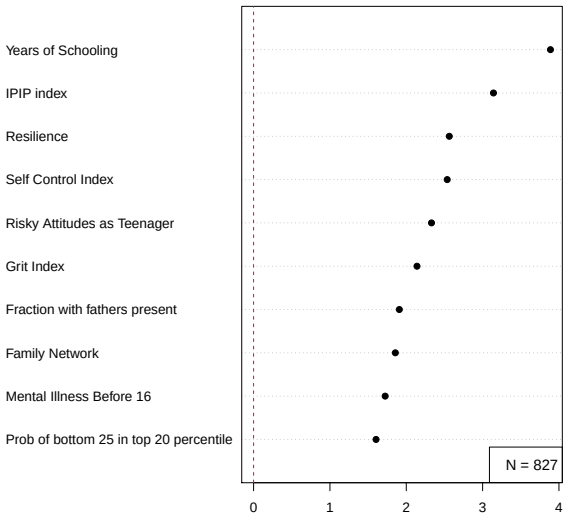


(b) Adjusted Income

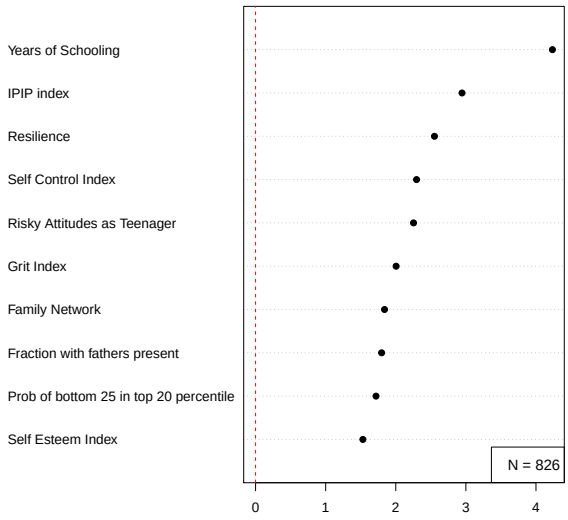


(c) Individual Income

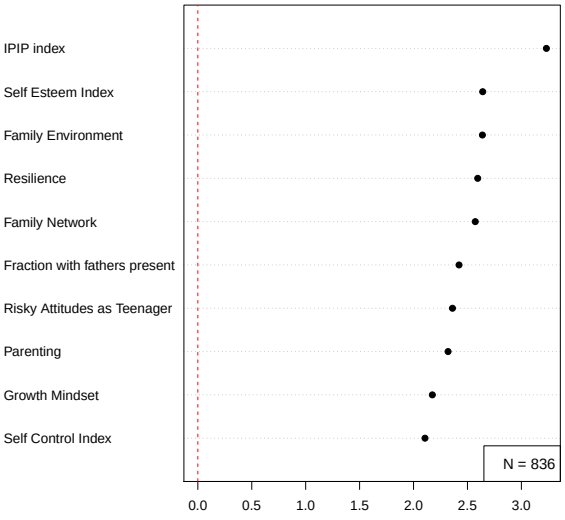
Figure 1: OLS Method



(a) Household Income

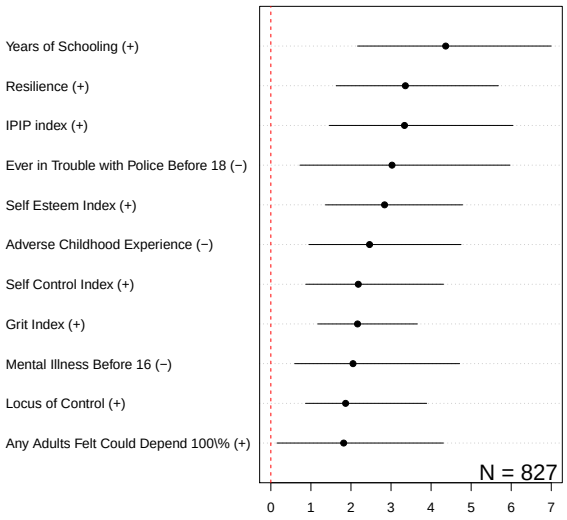


(b) Adjusted Income

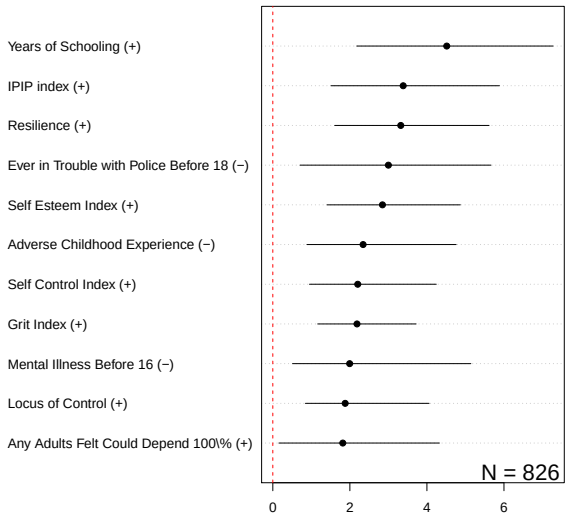


(c) Individual Income

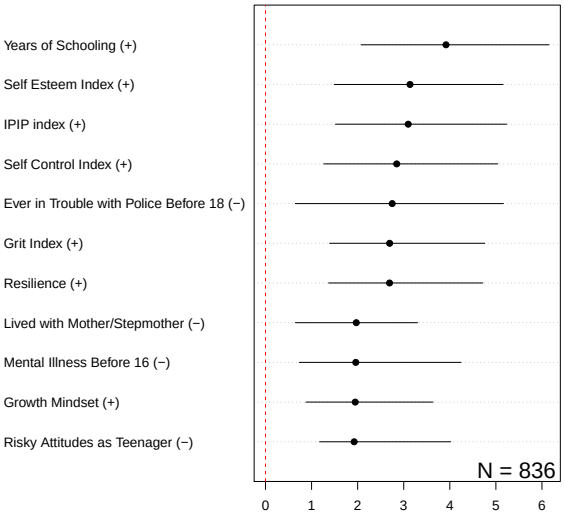
Figure 2: Top Variables - Random Forests



(a) Household Income

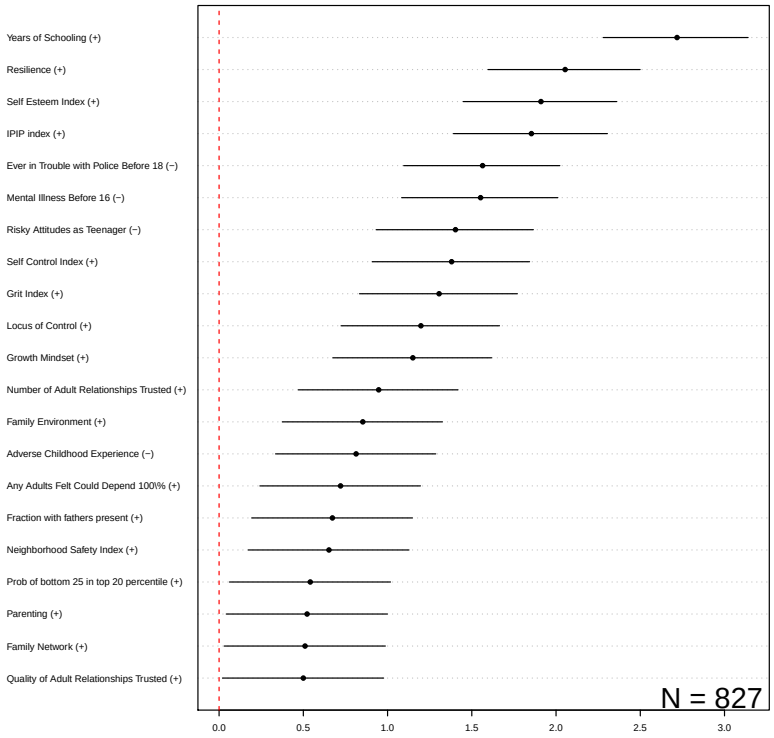


(b) Adjusted Income

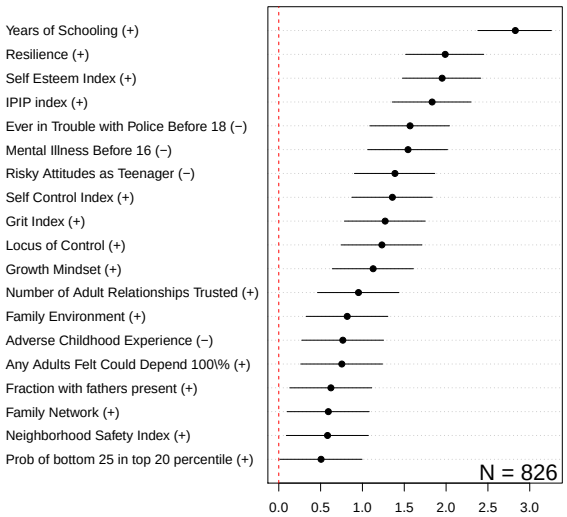


(c) Individual Income

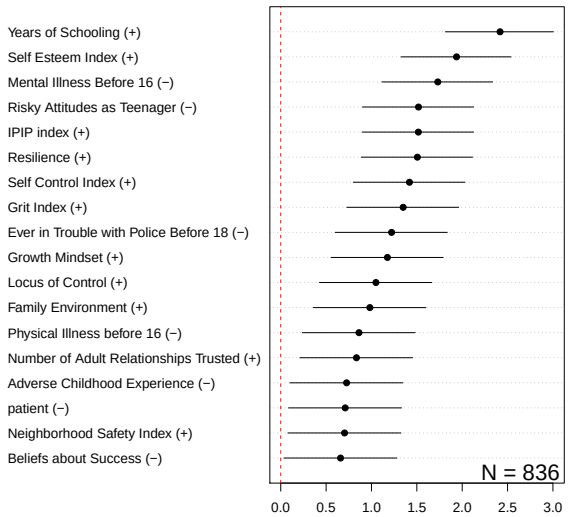
Figure 3: Non-Parametric Method



(a) Household Income

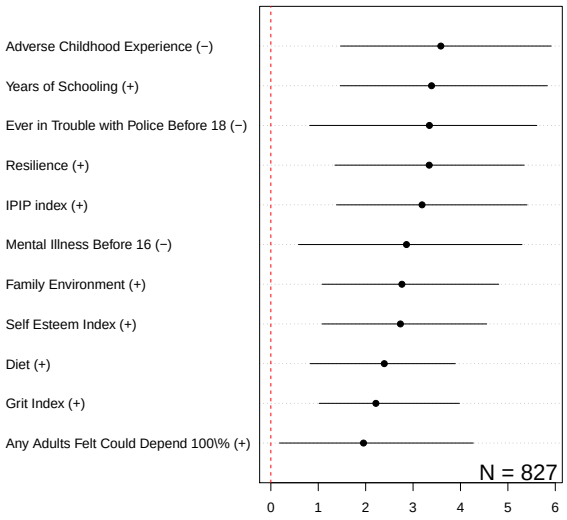


(b) Adjusted Income

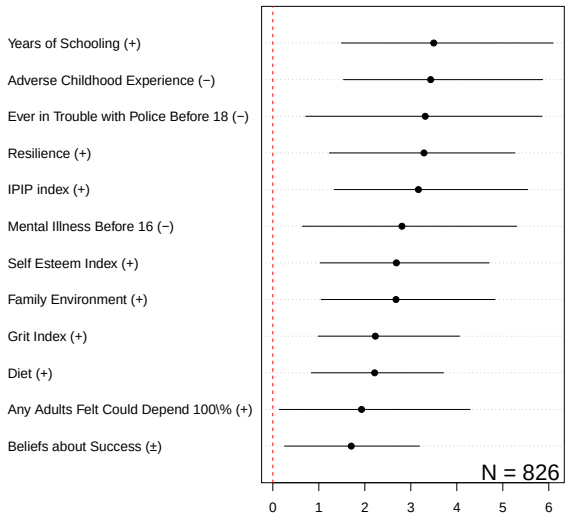


(c) Individual Income

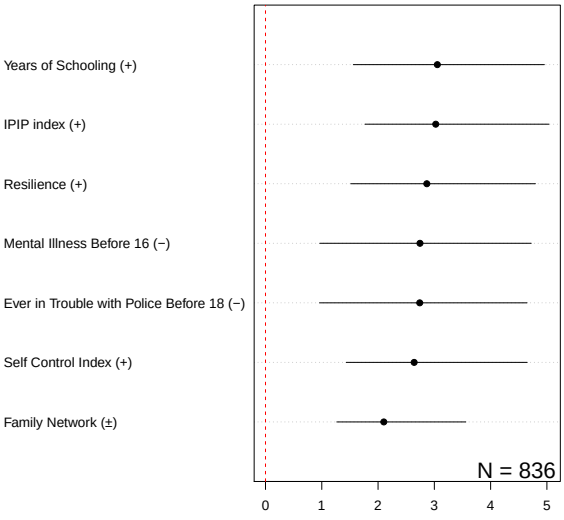
Figure 4: Partial Correlations



(a) Household Income

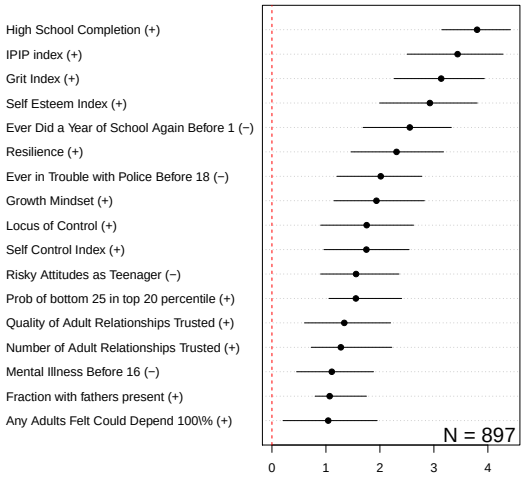


(b) Adjusted Income

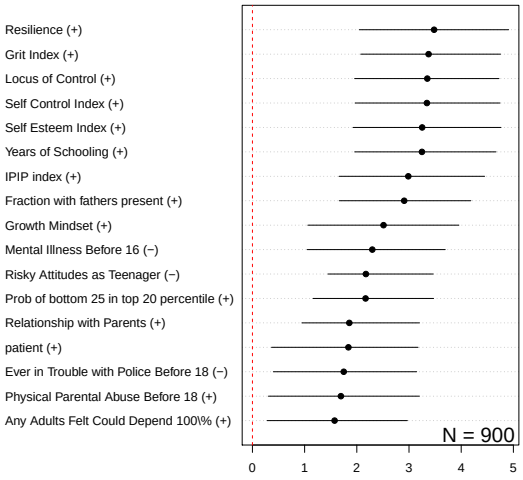


(c) Individual Income

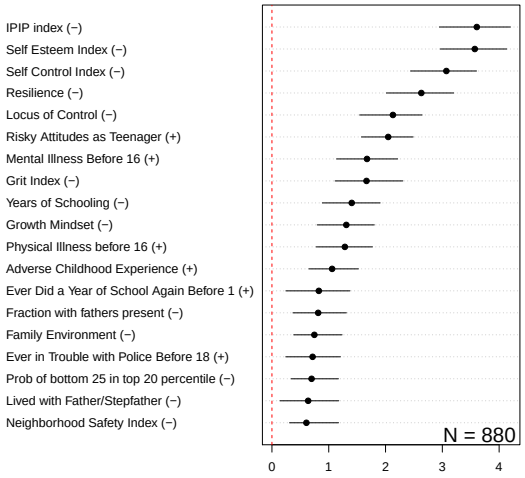
Figure 5: Nearest Neighbor



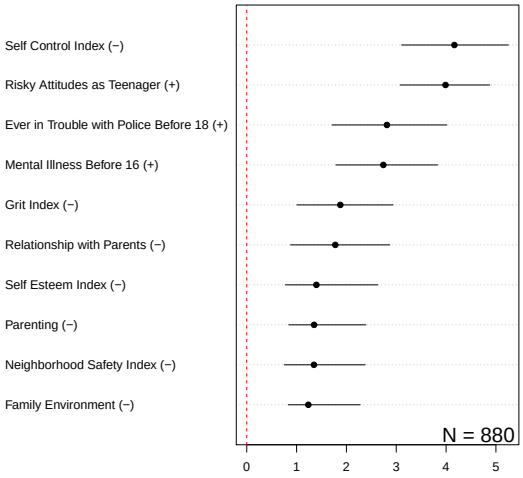
(a) College Graduation



(b) Married



(c) Adult Mental Illness



(d) Drug and Alcohol Use

Figure 6: Non-Parametric Method, Alternative Outcomes

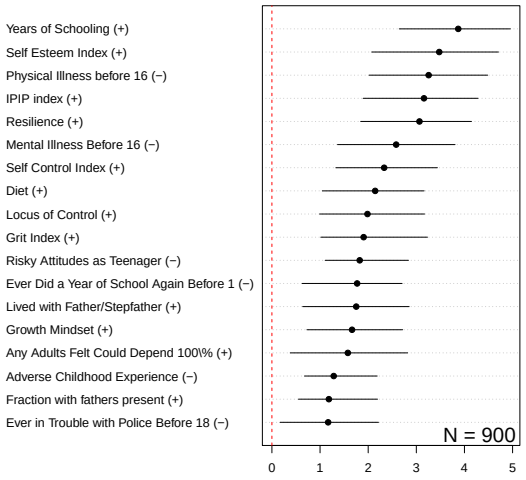


Figure 7: Adult Physical Health

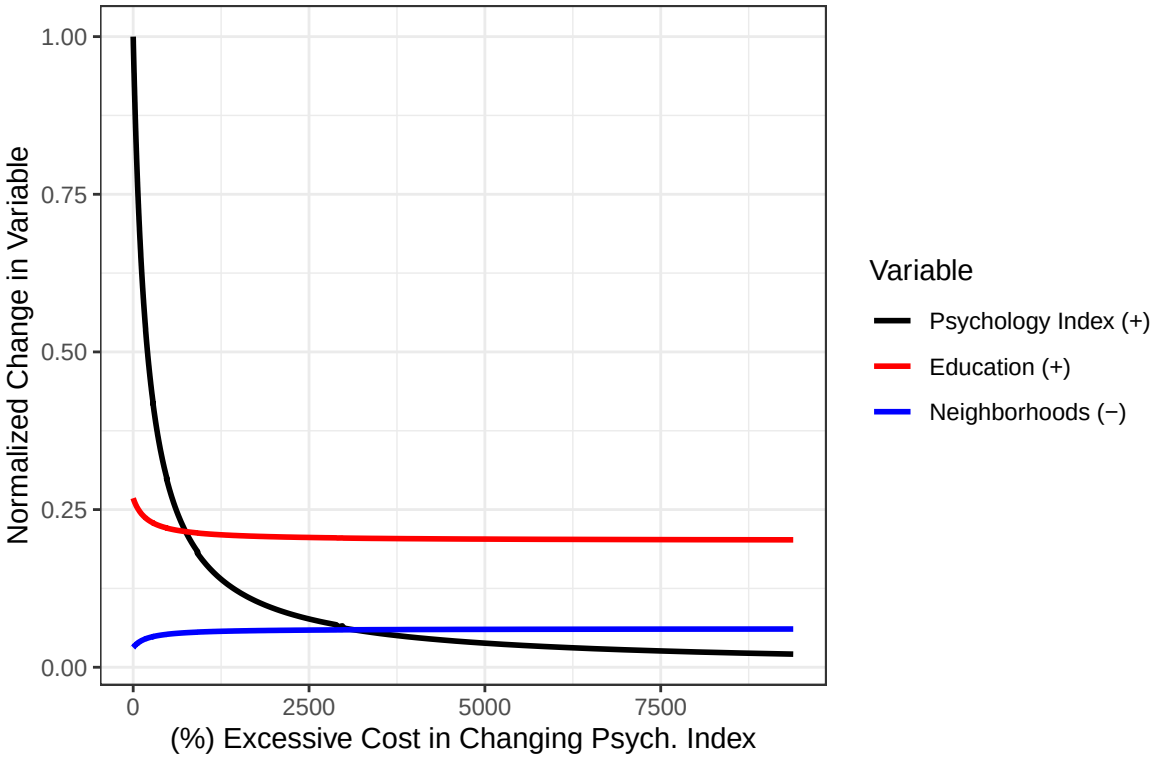
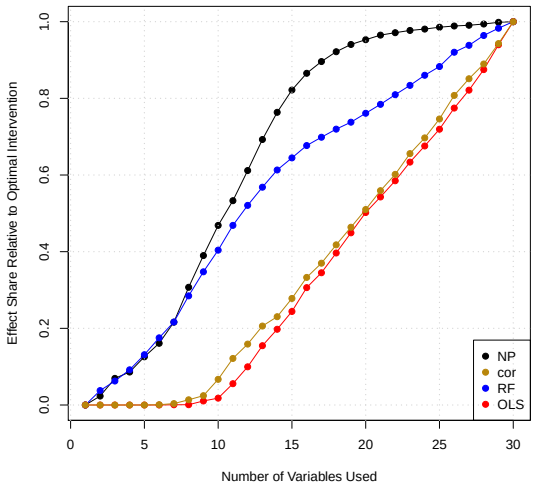
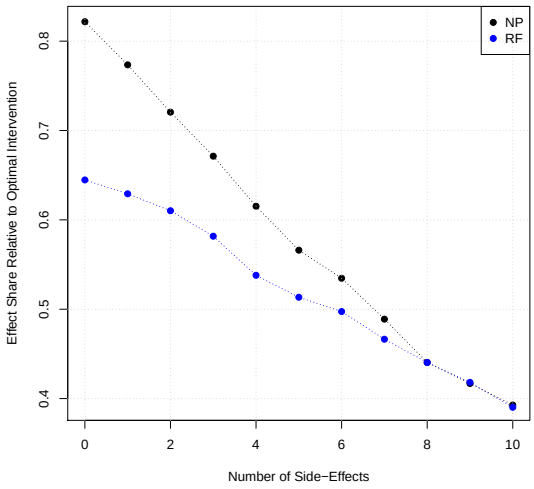


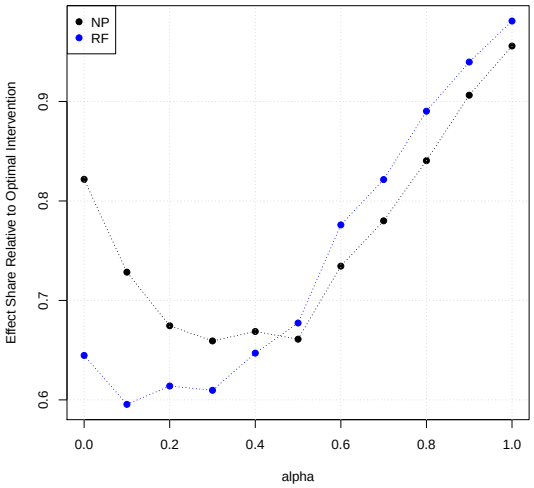
Figure 8: Excessive Cost



(a) Effect by Method



(b) Effect of Top 15 Variables With Side-Effects



(c) Effect of Top 15 Variables as Function of α

Figure 9: Simulations Results

References

[1] Aigner, and G. Cain. (1977).“Statistical Theories of Discrimination in Labor Markets." *Industrial and Labor Relations Review* 30, 175-187.

[2] Alaggia, R. (2010). An ecological analysis of child sexual abuse disclosure: Considerations for child and adolescent mental health. *Journal of the Canadian Academy of Child and Adolescent Psychiatry*, 19(1), 32.

[3] Aldaz-Carroll, E., & Moran, R. (2001). Escaping the poverty trap in Latin America: the role of family factors. *Cuadernos de economia*, 38(114), 155-190.

[4] Allgood, S., Mustard, D. B., & Warren Jr, R. S. (1999). The impact of youth criminal behavior on adult earnings. manuscript, University of Georgia.

[5] Almlund, M., Duckworth, A. L., Heckman, J., & Kautz, T. (2011). Personality psychology and economics. In *Handbook of the Economics of Education* (Vol. 4, pp. 1-181). Elsevier.

[6] Angrist, J. D., & Keueger, A. B. (1991). Does compulsory school attendance affect schooling and earnings?. *The Quarterly Journal of Economics*, 106(4), 979-1014.

[7] Baay, Pieter E., de Ridder, Denise T D, Eccles, Jacquelynne S., van der Lippe, T., van Aken, & Marcel A G. (2014). Self-control trumps work motivation in predicting job search behavior. *Journal of Vocational Behavior* 85(3), 443 - 451.

[8] Barnett, W. S., & Masse, L. N. (2007). Comparative benefit-cost analysis of the Abecedarian program and its policy implications. *Economics of Education Review*, 26(1), 113-125.

[9] Basu, D., Ghosh, A., Hazari, N., & Parakh, P. (2016). Use of Family CAGE-AID questionnaire to screen the family members for diagnosis of substance dependence. *The Indian journal of medical research*, 143(6), 722.

[10] Baum, S., Ma, J., & Payea, K. (2013). Education Pays, 2013: The Benefits of Higher Education for Individuals and Society. Trends in Higher Education Series. *College Board*.

[11] Becker, G. S., & Tomes, N. (1979). An equilibrium theory of the distribution of income and intergenerational mobility. *Journal of political Economy*, 87(6), 1153-1189.

[12] Befield, C. R., Nores, M., Barnett, S.W., & Schweinhart, L. (2006). Cost-benefit analysis of the High/Scope Perry Preschool Program using age 40 follow-up data. *Journal of Human Resources*, 41, 283-301

[13] Blackwell, L., Trzesniewski, K., & Dweck, C. (2007). Implicit Theories of Intelligence Predict Achievement across an Adolescent Transition: A Longitudinal Study and an Intervention. *Child Development*, 78, 246-263.

[14] Blau, P. M., & Duncan, O. D. (1967). The American occupational structure. *New York, London, Sidney: Wiley*.

[15] Brown, R. A., Monti, P. M., Myers, M. G., Martin, R. A., Rivinus, T., Dubreuil, M. E., & Rohsenow, D. J. (1998). Depression among cocaine abusers in treatment: relation to cocaine and alcohol use and treatment outcome. *American Journal of Psychiatry*, 155(2), 220-225.

[16] Brown, R. L., & Rounds, L. A. (1995). Conjoint screening questionnaires for alcohol and other drug abuse: criterion validity in a primary care practice. *Wisconsin medical journal*, 94(3), 135-140.

[17] Card, D. (1999). The causal effect of education on earnings. In *Handbook of labor economics* (Vol. 3, pp. 1801-1863). Elsevier.

[18] Chetty, R., & Hendren, N. (2018). The impacts of neighborhoods on intergenerational mobility I: Childhood exposure effects. *The Quarterly Journal of Economics*, 133(3), 1107-1162.

[19] Chetty, R., Hendren, N., & Katz, L. F. (2016). The effects of exposure to better neighborhoods on children: New evidence from the Moving to Opportunity experiment. *American Economic Review*, 106(4), 855-902.

[20] Chetty, R., Hendren, N., Jones, M. R., & Porter, S. R. (2018). *Race and economic opportunity in the United States: An intergenerational perspective* (No. w24441). National Bureau of Economic Research.

[21] 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.

[22] Claro, S., Paunesku, D., & Dweck, C. S. (2016). Growth mindset tempers the effects of poverty on academic achievement. *Proceedings of the National Academy of Sciences*, 113(31), 8664-8668.

[23] Currie, J., & Spatz Widom, C. (2010). Long-term consequences of child abuse and neglect on adult economic well-being. *Child maltreatment*, 15(2), 111-120.

[24] Darity, W. A., A. Goldsmith, & J. Veum (1997). The Impact of Psychological and Human Capital on Wages. *Economic Inquiry*. 35 (4), 818-829.

[25] De Ridder, D. T. D., Lensvelt-Mulders, G., Finkenauer, C., Stok, M., & Baumeister, R. F.(2012). Taking stock of self-control: A meta-analysis of how trait self-control relates to a wide range of behaviors. *Personality and Social Psychology Review*, 16, 76-99.

[26] Dinardo, J., Fortin, N., Lemieux, T., (1996). Labor Market Institutions and the Distribution of Wages, 1973-1992: A Semiparametric Approach, *Econometrica*

[27] Dobbie, W. S., & Fryer Jr, R. G. (2016). Charter schools and labor market outcomes (No. w22502). *National Bureau of Economic Research*.

[28] Dobbie, W., & Fryer Jr, R. G. (2013). Getting beneath the veil of effective schools: Evidence from New York City. *American Economic Journal: Applied Economics*, 5(4), 28-60.

[29] Dobbie, W., & Fryer Jr, R. G. (2015). The medium-term impacts of high-achieving charter schools. *Journal of Political Economy*, 123(5), 985-1037.

[30] Duckworth, A. L., & Quinn, P. D. (2009). Development and validation of the Short Grit Scale (GRIT-S). *Journal of personality assessment*, 91(2), 166-174.

[31] Duckworth, A.L., Peterson, C., Matthews, M.D., & Kelly, D.R., (2007). Grit: Perseverance and passion for long-term goals. *Journal of Personality and Social Psychology* 92, 1087-1101.

[32] Duckworth, A. L. & Seligman, M. E. P. (2005). Self-discipline outduels IQ in predicting academic performance of adolescents. *Psychological Science*, 16(12), 939-944.

[33] Duncan, G. J., & Rodgers, W. L. (1988). Longitudinal aspects of childhood poverty. *Journal of Marriage and the Family*, 1007-1021.

[34] Dweck, C. S. (1999). Essays in social psychology. *Self-theories: Their role in motivation, personality, and development*. New York, NY, US: Psychology Press.

[35] Elango, S., Garcia, J. L., Heckman, J. J., & Hojman, A. (2015). Early childhood education (No. w21766). *National Bureau of Economic Research*.

[36] Felitti, V. J., Anda, R. F., Nordenberg, D., Williamson, D. F., Spitz, A. M., Edwards, V., & Marks, J. S. (1998). Relationship of childhood abuse and household dysfunction to many of the leading causes of death in adults: The Adverse Childhood Experiences (ACE) Study. *American journal of preventive medicine*, 14(4), 245-258.

[37] Finkelstein, A., Taubman, S., Wright, B., Bernstein, M., Gruber, J., Newhouse, J. P., Allen, H., Baicker, K. & Oregon Health Study Group. (2012). The Oregon health insurance experiment: evidence from the first year. *The Quarterly journal of economics*, 127(3), 1057-1106.

[38] Fisher, R. A. (1925). *Statistical Methods for Research Workers*. Oliver & Boyd, Edinburgh.

[39] Freeman, R. B. (1986). Who Escapes?: The Relation of Churchgoing and Other Background Factors to the Socioeconomic Performance of Black Male Youths from Inner-City Poverty Tracts. *The Black Youth Employment Crisis*, pp. 353-376, Chicago: UCP.

[40] Fryer Jr, R. G. (2018). The “Pupil" Factory: Specialization and the Production of Human Capital in Schools. *American Economic Review*, 108(3), 616-56.

[41] Garces, E., Thomas, D., & Currie, J. (2002). Longer-term effects of Head Start. *American economic review*, 92(4), 999-1012.

[42] Goldberg, L. R. (1992). The development of markers for the Big-Five factor structure. *Psychological Assessment*, 4, 26-42.

[43] Goldberg, L. R., Johnson, J. A., Eber, H. W., Hogan, R., Ashton, M. C., Cloninger, C. R., & Gough, H. C. (2006). The International Personality Item Pool and the future of public-domain personality measures. *Journal of Research in Personality*, 40, 84-96.

[44] Gow, A. J., Whiteman, M. C., Pattie, A., & Deary, I. J. (2005). Goldberg’s IPIP Big-Five factor markers: Internal consistency and concurrent validation in Scotland. *Personality and Individual Differences*, 39(2), 317-329.

[45] Heckman, J. J., Humphries, J. E., & Veramendi, G. (2015). The causal effects of education on earnings and health. Unpublished manuscript, Department of Economics, University of Chicago.

[46] Heckman, J. J., Moon, S. H., Pinto, R., Savelyev, P. A., & Yavitz, A. (2010). The rate of return to the HighScope Perry Preschool Program. *Journal of public Economics*, 94(1-2), 114-128.

[47] Heckman, J. J., Stixrud, J., & Urzua, S. (2006). The effects of cognitive and noncognitive abilities on labor market outcomes and social behavior. *Journal of Labor economics*, 24(3), 411-482.

[48] Heckman, James J., and Yona Rubinstein. (2001).“The importance of noncognitive skills: Lessons from the GED testing program.’ *American Economic Review* 145-149.

[49] Heckman, J., Pinto, R., & Savelyev, P. (2013). Understanding the mechanisms through which an influential early childhood program boosted adult outcomes. *American Economic Review*, 103(6), 2052-86.

[50] Isaacs, J. B., Sawhill, I. V., & Haskins, R. (2008). Getting Ahead or Losing Ground: Economic Mobility in America. *Brookings Institution*.

[51] Keels, M., Duncan, G. J., DeLuca, S., Mendenhall, R., & Rosenbaum, J. (2005). Fifteen years later: Can residential mobility programs provide a long-term escape from neighborhood segregation, crime, and poverty. *Demography*, 42(1), 51-73.

[52] Kroenke, K., Spitzer, R. L., & Williams, J. B. (2001). The PHQ-9: validity of a brief depression severity measure. *Journal of general internal medicine*, 16(9), 606-613.

[53] Lindner, C., Nagy, G., & Retelsdorf, J. (2015). The dimensionality of the Brief Self-Control Scale - An evaluation of unidimensional and multidimensional applications. *Personality and Individual Differences*, 86, 465-473.

[54] Lowe, B., Kroenke, K., Herzog, W., & Grafe, K. (2004). Measuring depression outcome with a brief self-report instrument: sensitivity to change of the Patient Health Questionnaire (PHQ-9). *Journal of affective disorders*, 81(1), 61-66.

[55] McKernan, S. M., & Ratcliffe, C. (2005). Events that trigger poverty entries and exits. *Social Science Quarterly*, 86, 1146-1169.

[56] Metzler, M., Merrick, M. T., Klevens, J., Ports, K. A., & Ford, D. C. (2017). Adverse childhood experiences and life opportunities: shifting the narrative. *Children and youth services review*, 72, 141-149.

[57] Moffitt, T. E., Arseneault, L., Belsky, D., Dickson, N., Hancox, R. J., Harrington, H., H., Houts, R., Poulton, R., Roberts, B.W., Ross, S. & Sears, M. R. (2011). A gradient of childhood self-control predicts health, wealth, and public safety. *Proceedings of the National Academy of Sciences*, 108(7), 2693-2698.

[58] Mullainathan, S., & Spiess, J. (2017). Machine learning: an applied econometric approach. *Journal of Economic Perspectives*, 31(2), 87-106.

[59] Nyhus, E. K., & Pons, E. (2005). The effects of personality on earnings. *Journal of Economic Psychology*, 26(3), 363-384.

[60] OECD. (2018). A Broken Social Elevator? How to Promote Social Mobility. *OECD Publishing, Paris*.

[61] Rosenberg, M. (1965). Society and the Adolescent Self-Image. *Princeton, NJ: Princeton University Press*.

[62] Rotter J. (1966): Generalized Expectancies for Internal versus External Control of Reinforcement, *Psychological Monographs*, 80, 1-28.

[63] Sanbonmatsu, L., Potter, N. A., Adam, E., Duncan, G. J., Katz, L. F., Kessler, R. C., Ludwig, J., Marvakov, J., Yang, F., Congdon, W.J. & Gennetian, L. A. (2012). The long-term effects of moving to opportunity on adult health and economic self-sufficiency. *Cityscape*, 109-136.

[64] Sharkey, P., & Torrats-Espinosa, G. (2017). The effect of violent crime on economic mobility. *Journal of Urban Economics*, 102, 22-33.

[65] Smith, B. W., Dalen, J., Wiggins, K., Tooley, E., Christopher, P., & Bernard, J. (2008). The brief resilience scale: assessing the ability to bounce back. *International journal of behavioral medicine*, 15(3), 194-200.

[66] Solon, Gary & Marianne E. Page & Greg J. Duncan. (2000). "Correlations Between Neighboring Children In Their Subsequent Educational Attainment," *The Review of Economics and Statistics*, MIT Press, 82(3), 383-392.

[67] Solon, G. (1992). Intergenerational income mobility in the United States. *The American Economic Review*, 393-408.

[68] Spitzer, R. L., Kroenke, K., Williams, J. B., & Lowe, B. (2006). A brief measure for assessing generalized anxiety disorder: the GAD-7. *Archives of internal medicine*, 166(10), 1092-1097.

[69] Spitzer, R. L., Williams, J. B., Kroenke, K., Hornyak, R., McMurray, J., & Patient Health Questionnaire Obstetrics-Gynecology Study Group. (1999). Validity and utility of the PRIME-MD patient health questionnaire in assessment of 3000 obstetric-gynecologic patients: the PRIME-MD Patient Health Questionnaire Obstetrics-Gynecology Study. *American journal of obstetrics and gynecology*, 183(3), 759-769.

[70] Strayhorn, J. M., & Weidman, C. S. (1988). A parent practices scale and its relation to parent and child mental health. *Journal of the American Academy of Child & Adolescent Psychiatry*, 27(5), 613-618.

[71] Tangney, J.P., Baumeister, R.F., Boone, A.L. (2004). High self-control predicts good adjustment, less pathology, better grades, and interpersonal success. *Journal of Personality*, 72, 271-322.

[72] Western, B., & Pettit, B. (2010). Incarceration & social inequality. *Daedalus*, 139(3), 8-19.

[73] Windle, G., Bennett, K. M., & Noyes, J. (2011). A methodological review of resilience measurement scales. *Health and quality of life outcomes*, 9(1), 8.

Appendix Figures

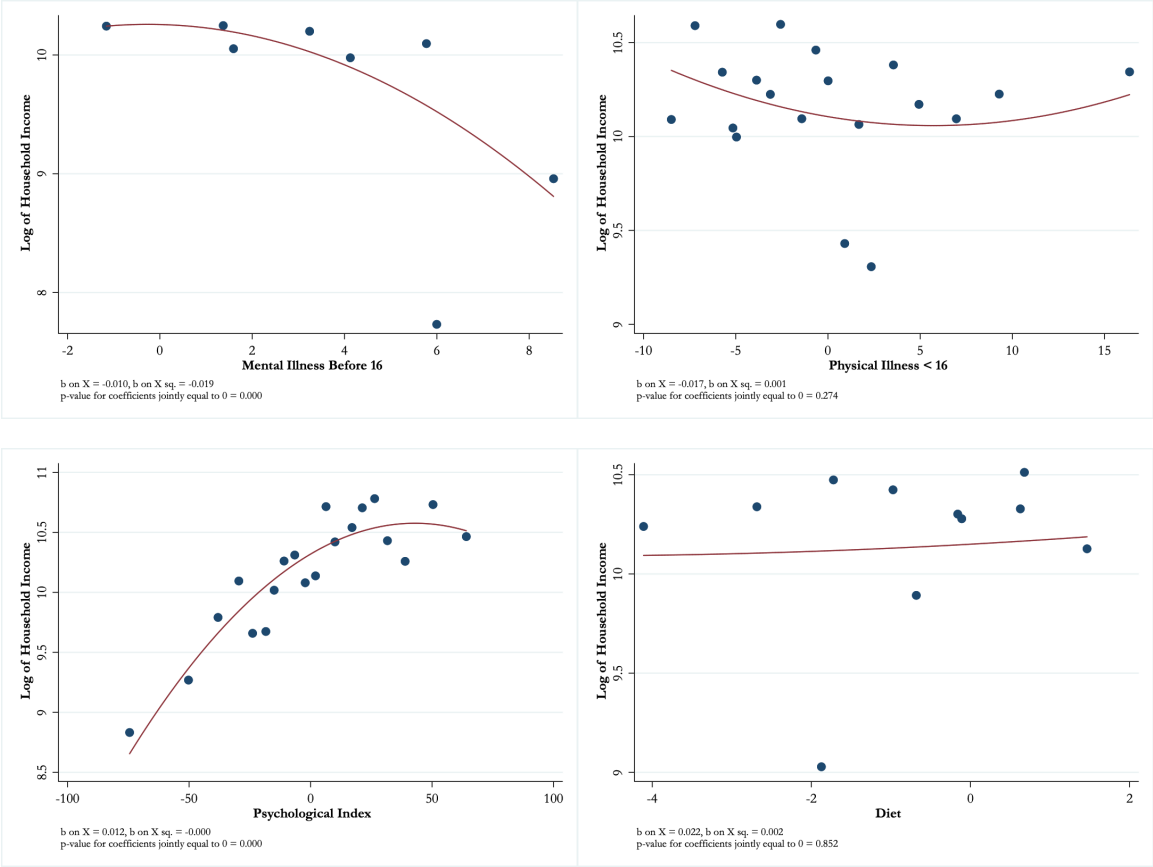


Figure 10: Bin Scatter of Household Income with Health Indices

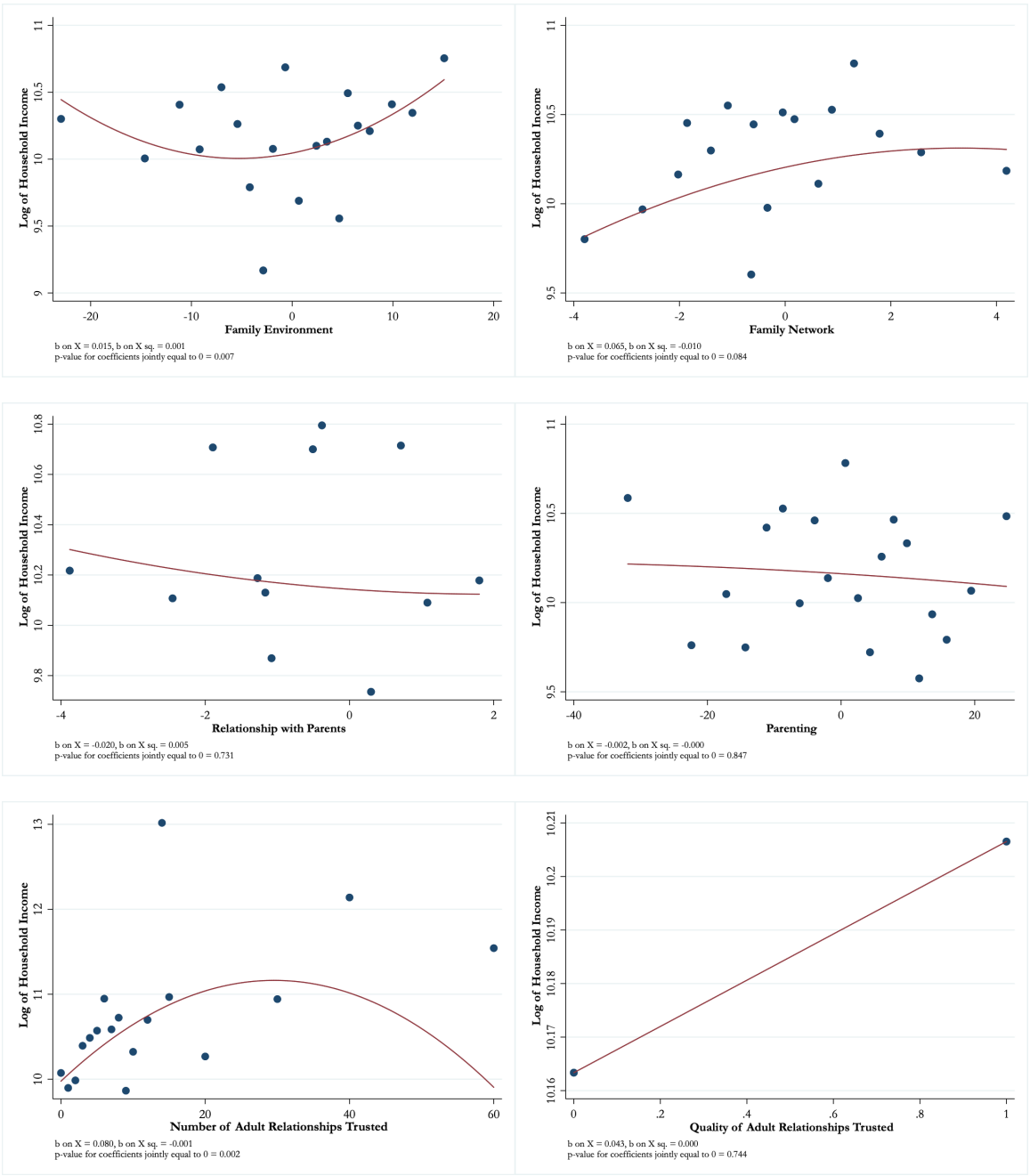


Figure 11: Bin Scatter of Household Income with Family Based Indices

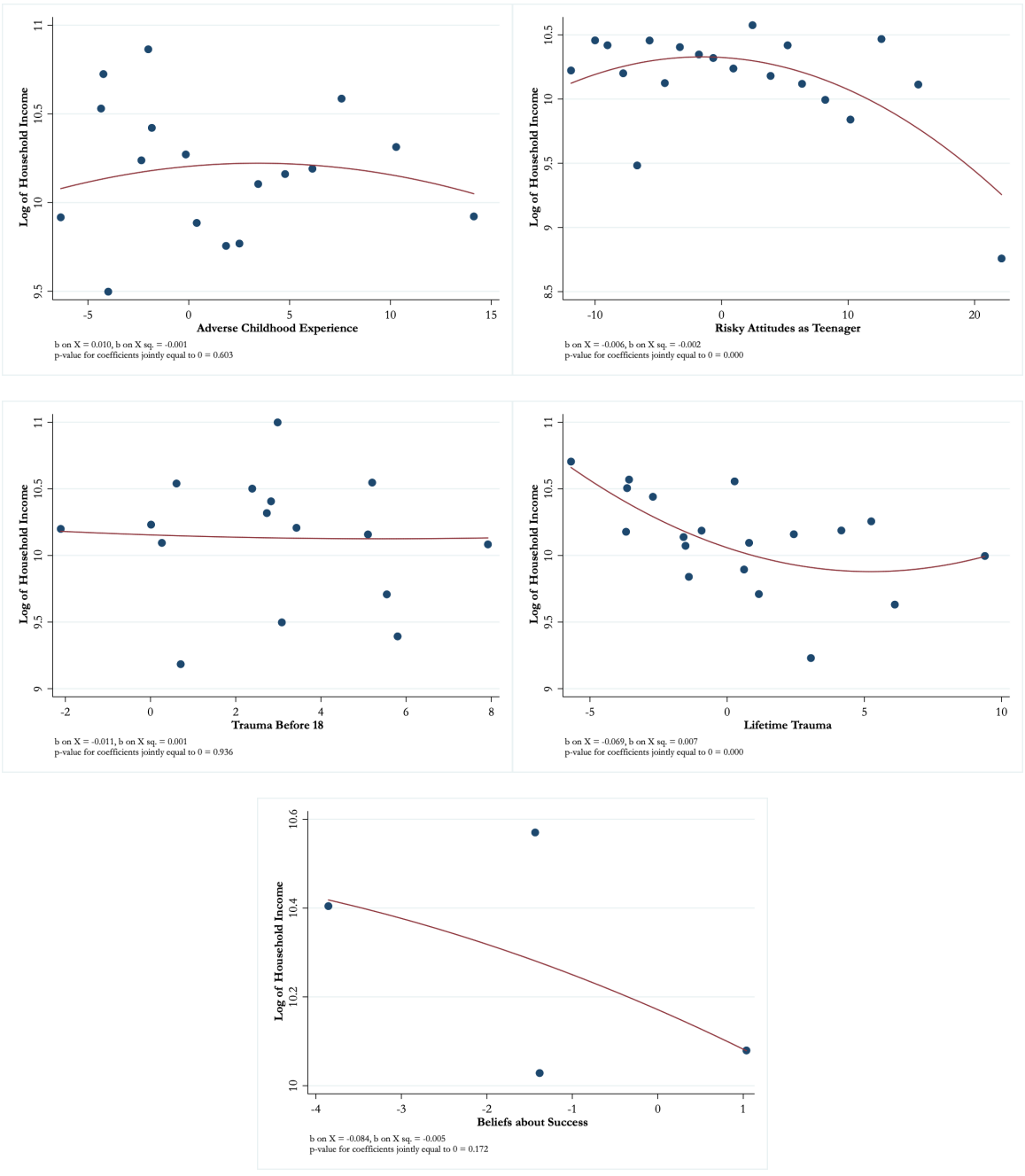


Figure 12: Bin Scatter of Household Income with Childhood Experiences Indices

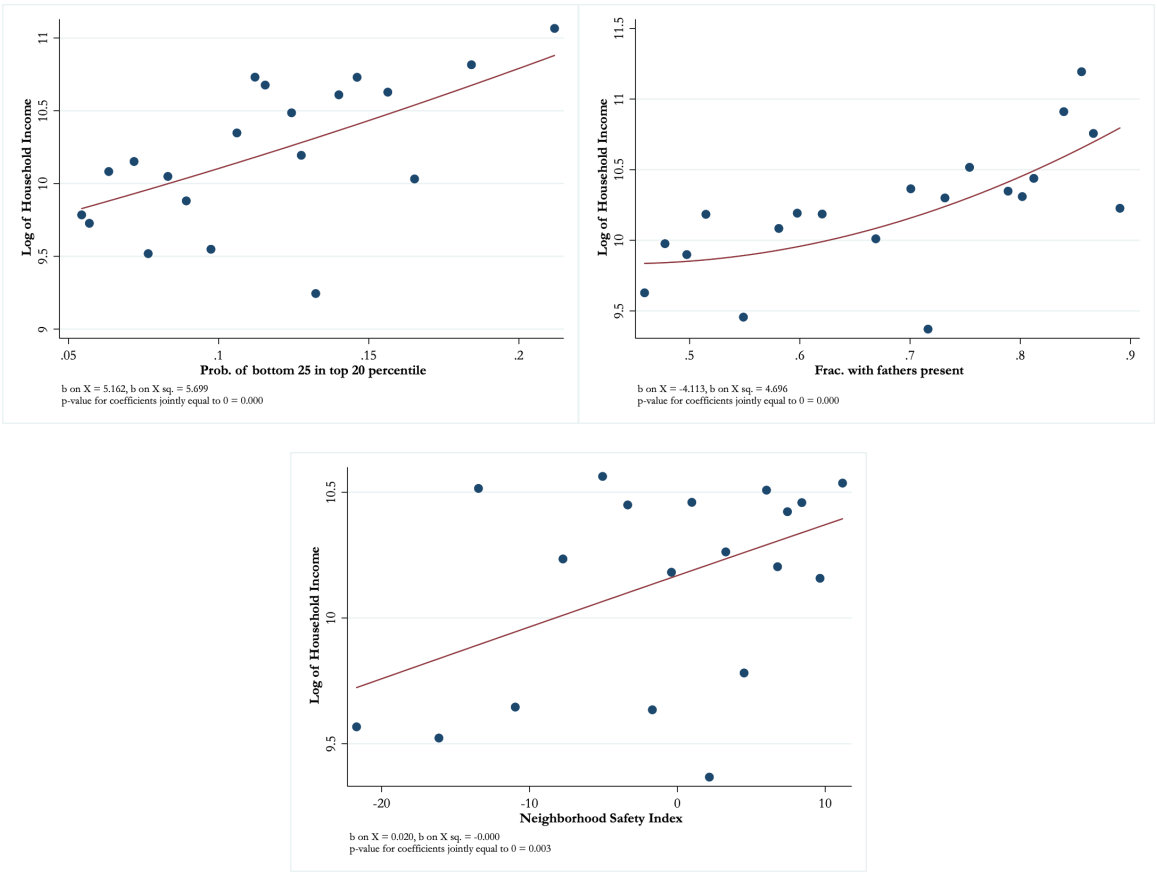
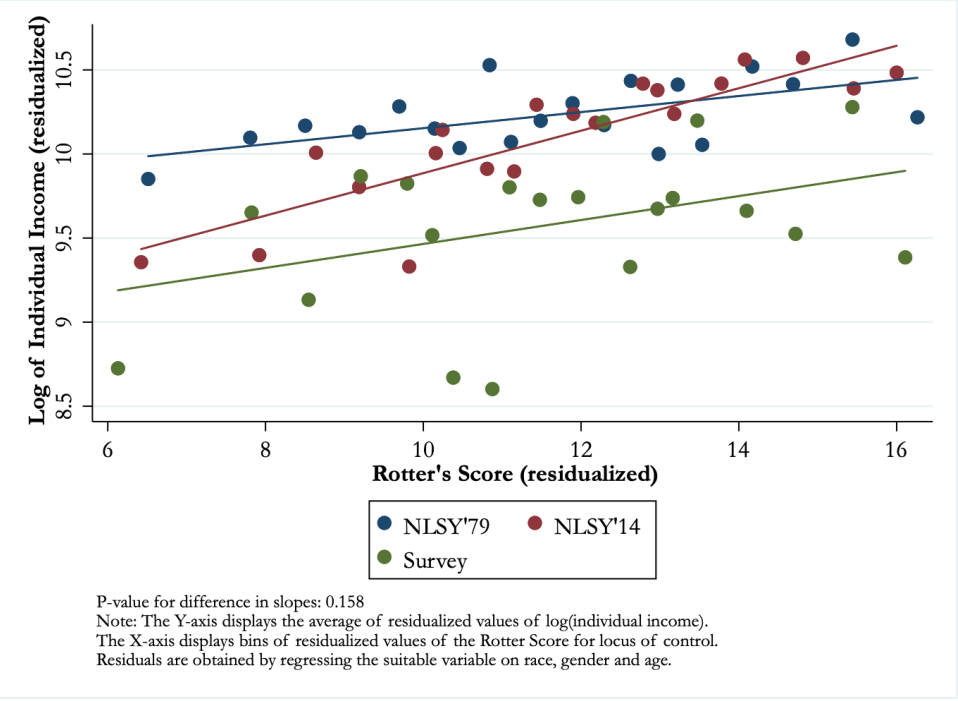


Figure 13: Bin Scatter of Household Income with Neighborhood Indices



P-value for difference in slopes: 0.158
Note: The Y-axis displays the average of residualized values of log(individual income).
The X-axis displays bins of residualized values of the Rotter Score for locus of control.
Residuals are obtained by regressing the suitable variable on race, gender and age.

Figure 17: Correlation Between Income and Locus of Control

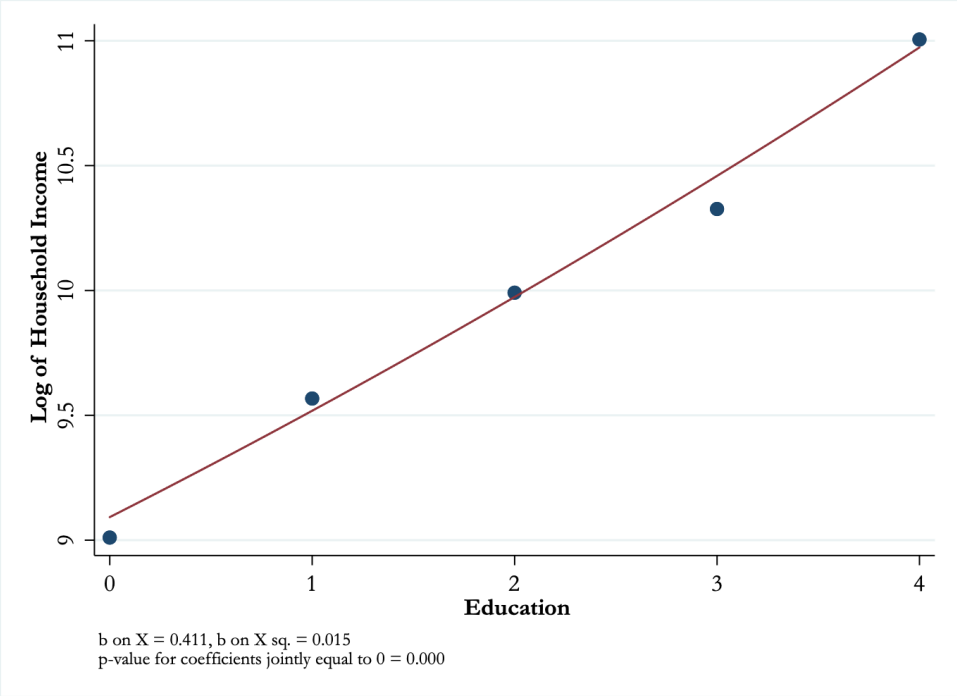
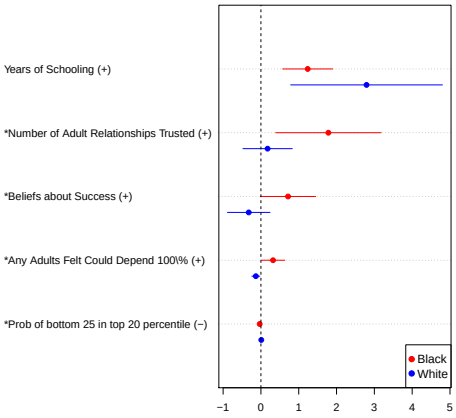
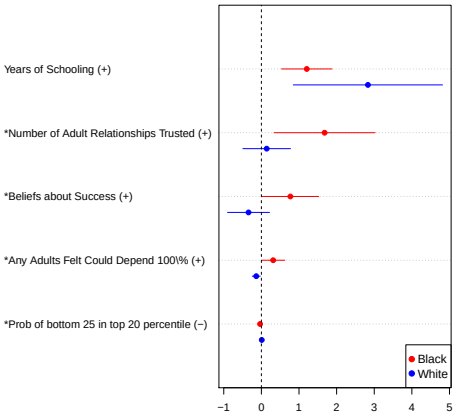


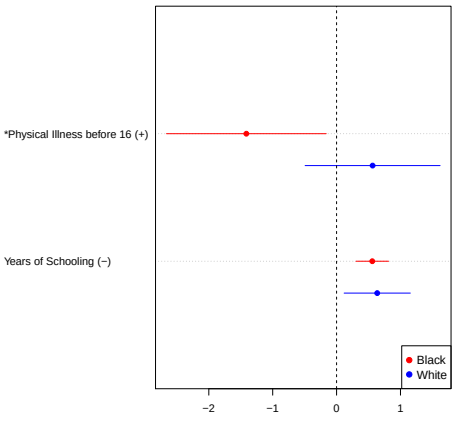
Figure 14: Bin Scatter of Household Income with Education



(a) Household Income

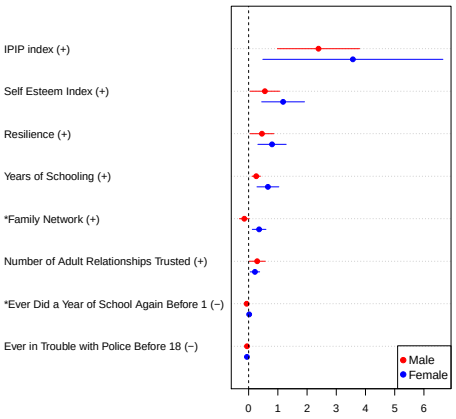


(b) Adjusted Income

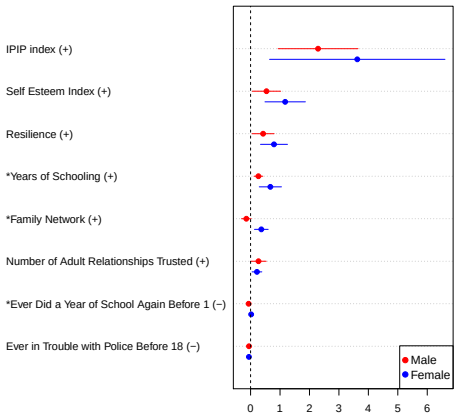


(c) Individual Income

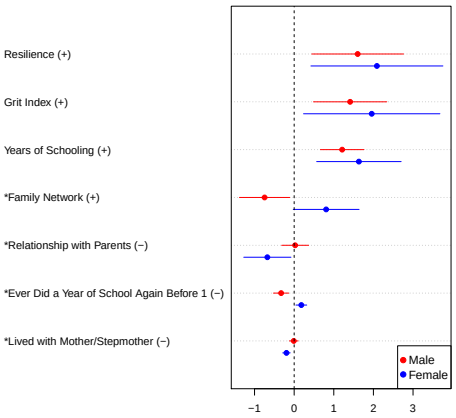
Figure 15: Subsample Analysis by Race



(a) Household Income



(b) Adjusted Income



(c) Individual Income

Figure 16: Subsample Analysis by Gender

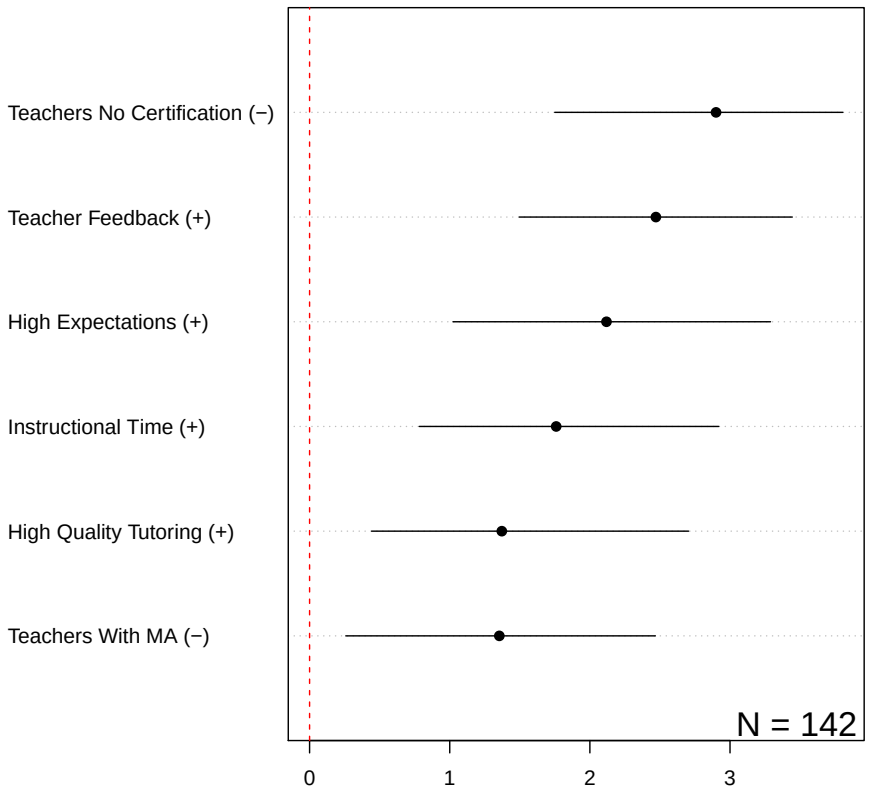


Figure 18: Non-Parametric Correlations, Dobbie and Fryer (2013)

Appendix Tables

		Unadjusted	Adjusted	
	Survey	NLSY	NLSY	*p-value*
	(1)	(2)	(3)	(1)=(2)
Panel A: Demographics				
Top-coded Individual Income	28,311.881	27,751.124	24,246.806	0.759
Age	48.689	52.579	52.729	0.000
Female	0.527	0.513	0.520	0.618
Black	0.432	0.384	0.435	0.066
Hispanic	0.096	0.145	0.147	0.001
Panel B: Adult Mental Illness Sub-Categories				
Poor appetite or Overeating	1.859	1.520	1.512	0.000
Trouble concentrating on things	1.562	1.588	1.543	0.638
Feeling depressed	1.656	1.514	1.488	0.005
Trouble falling asleep	2.184	1.997	1.935	0.004
Panel C: Psychological Traits				
Rotter Locus of Control Score	11.819	11.992	11.825	0.252
Rosenberg Self Esteem Score	21.858	23.101	23.100	0.000
Observations	900	2,534		

Table 2: Summary Statistics by Survey

Notes: This table presents summary statistics for different survey samples. Column (1) presents weighted averages for the authors' survey in Shelby County, TN, Tulsa County, OK, and Jefferson and Orleans Parishes, LA. Columns (2) and (3) present weighted averages for respondents in the National Longitudinal Survey of Youth (NLSY) 1979 who were classified as poor in 1979. Column (4) presents the p-value for the difference in values between columns (1) and (2). Column (5) presents the p-value for the difference in values between columns (1) and (3). Weights for column (1) are sampling weights taken from the survey. Weights for column (2) are sampling weights taken from the 2014 wave of NLSY'79. Weights for column (3) were calculated to age-gender-race adjust the NLSY sample to look like the survey sample. Age groups are (a) 49-51 (b) 52-54 and (c) 55-57. Gender groups are (a) male and (b) female. Race groups are (a) Black, (b) Hispanic and (c) Other. Individual income for the NLSY sample was taken from its most recent 2014 wave. Individual incomes from the survey were top coded using the same rules as the NLSY sample. Age, gender and race information for the NLSY sample was taken from its 1978 screener information. For adult mental illness sub-categories, NLSY respondents were asked these questions once they turned 40 and then once when they turned 50. We use the response from when they turned 50. If that was missing and there was a non-missing response from when they turned 40, we used their previous response. All adult mental illness reponses were categorical where 1 = Rarely/None of the time/1 Day, 2 = Some/A little of the time/1-2 Days, 3 = Occasionally/Moderate amount of the time/3-4 Days, and 4 = Most/All of the time/5-7 Days. For Rotter Locus of Control, we use NLSY's 2014 measure. The Rotter score for our survey and the NLSY sample is a score which is between 4-16 where a higher score indicates more internal locus of control. For Rosenberg Self Esteem, we use NLSY's 2006 measure. Rosenberg self esteem score for our survey and the NLSY sample is a score between 0-30 where a higher score indicates higher self esteem. All variables are explained in detail in the Data Appendix.

	Survey	Census	*p-value*
	(1)	(2)	(3)
Panel A: Demographics			
Household Income	47,483.895	91,850.000	0.000
Female	0.527	0.508	0.369
Black	0.437	0.127	0.000
Hispanic	0.097	0.178	0.000
High School/GED/Incomplete College	0.488	0.407	0.000
Two-year Associate Degree	0.119	0.064	0.000
Bachelor or Some Post-Graduate Degree	0.208	0.231	0.170
Observations	900	3,156,487	

Table 3: Summary Statistics (Survey vs. National)

Notes: This table presents summary statistics for different survey samples. Column (1) presents weighted averages for the authors' survey in Shelby County, TN, Tulsa County, OK, and Jefferson and Orleans Parishes, LA. Column (2) presents weighted averages for participants in the American Community Survey 2016 (ACS 2016). Column (3) presents the p-value for the difference in values between columns (1) and (2). Weights for column (1) are sampling weights taken from the survey. Weights for column (2) are sampling weights taken from ACS 2016 . All variables are explained in detail in the Data Appendix.

	Phone	Non-Eligible	Eligible	Agreed to	Paper	*p-value*	*p-value*	*p-value*
	Survey			Participate	Survey			
	(1)	(2)	(3)	(4)	(5)	(2)=(3)	(3)=(4)	(3)=(5)
Panel A: Demographics								
Poor Now?	0.212	0.095	0.385	0.419	0.413	0.000	0.001	0.029
Age	52.982	53.522	52.191	50.865	50.990	0.005	0.001	0.012
Female	0.553	0.571	0.527	0.526	0.545	0.000	0.932	0.155
Panel B: Race								
White	0.472	0.572	0.326	0.310	0.310	0.000	0.094	0.186
Black	0.226	0.198	0.266	0.334	0.357	0.000	0.000	0.000
Hispanic	0.251	0.181	0.355	0.291	0.271	0.000	0.000	0.000
Panel C: Education								
Less than HS	0.073	0.023	0.145	0.120	0.102	0.000	0.001	0.000
Some HS, Incomplete	0.074	0.051	0.108	0.109	0.105	0.000	0.926	0.671
HS/GED/Incomplete Col.	0.396	0.374	0.427	0.434	0.424	0.000	0.478	0.820
Two Year Associate	0.099	0.103	0.092	0.103	0.117	0.137	0.068	0.002
Bachelor or Graduate	0.183	0.224	0.124	0.124	0.134	0.000	0.914	0.243
Observations	6,459	3,842	2,617	1,227	928			

Table 4: Summary Statistics by Sample

Notes: This table presents summary statistics for different samples. Column (1) presents averages for all participants with completed phone screeners. Column (2) presents averages for all participants who were deemed non-eligible to participate in the survey. Column (3) presents averages for all participants who were deemed eligible to participate in the survey. Column (4) presents averages for all participants who were deemed eligible and who agreed to participate in the paper survey. Column (5) presents averages for all participants who participated in the paper survey. Column (6) presents the p-value for the difference in values between people who were deemed non-eligible and people who were deemed eligible for the paper survey. Column (7) presents the p-value for the difference in values between people who were deemed eligible but did not agree and people who were deemed eligible and agreed to participate in the survey. Column (8) presents the p-value for the difference in values between people who finally participated in the paper survey and people who were deemed eligible but did not participate in the paper survey.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)	(19)	(20)
(1)	1.00																			
(2)	-0.06	1.00																		
(3)	-0.01	0.32	1.00																	
(4)	0.31	-0.23	-0.12	1.00																
(5)	0.03	-0.13	-0.18	0.10	1.00															
(6)	-0.05	-0.29	-0.14	0.21	0.22	1.00														
(7)	0.00	0.01	-0.00	0.05	0.16	0.11	1.00													
(8)	-0.03	-0.28	-0.14	0.10	0.17	0.30	0.10	1.00												
(9)	-0.01	-0.28	-0.17	0.17	0.38	0.48	0.16	0.51	1.00											
(10)	0.04	-0.16	-0.07	0.13	0.17	0.22	0.11	0.22	0.29	1.00										
(11)	0.10	-0.13	0.00	0.13	0.16	0.18	0.09	0.09	0.24	0.37	1.00									
(12)	0.12	-0.13	-0.01	0.13	0.09	0.20	0.04	0.24	0.31	.	0.14	1.00								
(13)	-0.00	0.42	0.31	-0.15	-0.32	-0.43	-0.10	-0.46	-0.57	-0.21	-0.14	-0.24	1.00							
(14)	-0.06	0.44	0.32	-0.33	-0.08	-0.31	0.04	-0.20	-0.34	-0.05	-0.03	-0.08	0.36	1.00						
(15)	-0.11	0.31	0.32	-0.09	-0.13	-0.27	-0.05	-0.26	-0.33	-0.07	-0.07	-0.13	0.51	0.41	1.00					
(16)	-0.12	0.23	0.28	-0.16	-0.11	-0.13	0.04	-0.12	-0.16	-0.06	-0.04	-0.08	0.34	0.30	0.32	1.00				
(17)	0.01	-0.05	-0.01	0.154	0.03	0.09	0.04	0.08	0.06	0.07	0.04	0.01	-0.02	0.03	0.02	-0.06	1.00			
(18)	0.18	-0.03	-0.01	0.11	0.08	-0.00	-0.01	-0.00	-0.00	0.04	0.06	0.06	-0.04	-0.03	-0.04	-0.01	-0.05	1.00		
(19)	0.14	-0.00	0.01	0.14	0.03	0.003	-0.03	0.01	-0.04	0.03	0.07	0.09	-0.00	-0.03	0.00	-0.01	-0.06	0.77	1.00	
(20)	0.00	-0.22	-0.14	0.12	0.17	0.232	0.06	0.16	0.22	0.12	0.10	0.03	-0.31	-0.31	-0.22	-0.15	-0.08	0.10	0.15	1.00

Table 5: Correlation Matrix

Notes: This table presents pairwise correlations between all pairs of variables. The variables considered are – (1) Education, (2) Mental Illness Before 16, (3) Physical Illness Before 16, (4) Psychological Index, (5) Diet, (6) Family Environment, (7) Family Network, (8) Relationship with Parents, (9) Parenting, (10) Trust any Adults in Childhood, (11) Number of Adult Relationships Trusted, (12) Quality of Adult Relationships Trusted, (13) Adverse Childhood Experience, (14) Risky Attitudes as Teenager, (15) Trauma Before 18, (16) Lifetime Trauma, (17) Beliefs about Success, (18) Prob. of bottom 25 in top 20 percentile, (19) Frac. with fathers present, (20) Neighborhood Safety Index. Note that indices for lifetime trauma before 18 years of age and any lifetime trauma and the psychological index were created from sub-indices for this matrix only. Lifetime trauma before 18 Index was created out of 4 individual questions. Any lifetime trauma was created using 9 individual questions. Psychological index was created as a sum of 7 sub-indices – grit, resilience, IPIP, self esteem, self control, locus of control, and growth mindset. All variables are explained in detail in the Appendix D.3.

A Proofs

A.1 Formal Definition of Axioms

A1: Continuity: Let $\{P_0^n\}_{n=1}^{\infty}, \{P_I^n\}_{n=1}^{\infty}$ be a series of distributions over X that satify the regularity conditions of Theorem 1 (see below) with densities $\{\pi_0^n\}_{n=1}^{\infty}, \{\pi_I^n\}_{n=1}^{\infty}$ such that $\frac{\pi_0^n}{\pi_0}, \frac{\pi_I^n}{\pi_I} \to 1$ uniformly, then

$$C\left(P_I^n|P_0^n\right) \to C\left(P_I|P_0\right)$$

A2: Weakly positive: $C\left(P_I|P_0\right) \geq 0$, where $C\left(P_I|P_0\right) = 0$ iff $P_I = P_0$ with probability 1.

A3: Invariance of representation: Let P_0, P_I be some distributions on X. Define an invertible function $\Phi : X \to Z$. Let Q_0, Q_I be distributions on Z s.t. $P_i\left(X\right) = Q_i\left(\Phi\left(X\right)\right)$ for $i = 0, I$. Then $C\left(P_I|P_0\right) = C\left(Q_I|Q_0\right)$

A4: Additivity: If Xcan be written as $X = X_1 \times X_2$, define $\mathcal{I}\left(X_1\right)$ as the set of interventions for which the conditional distribution of X_2 is unchanged ($P_I\left(X_2|X_1\right) = P_0\left(X_2|X_1\right)$), and let $\mathcal{I}^{\perp}\left(X_1\right)$ be the set of interventions that are keeping X_1 unaffected ($P_I\left(X_1\right) = P_0\left(X_1\right)$). Then for any combination of $I_1 \in \mathcal{I}\left(X_1\right), I_2 \in \mathcal{I}^{\perp}\left(X_1\right)$, if $I = I_1 + I_2$ (combining both interventions) -

$$C\left(P_{I_{1+2}}|P_0\right) = C\left(P_{I_1}|P_{I_0}\right) + E_{X_1}\left[C\left(P_{I_2}|P_{I_1}\right)\right]$$

A.2 Proof of Theorem 1

We will first repeat the theorem with all the regularity conditions

Theorem. *Let P_I, P_0 be functions that satisfy:*

1. *P_0, P_I has densities π_0, π_I that are Lipschitz continuous with probability 1.*[23]
2. *$0 < \frac{\pi_I}{\pi_0} < \infty$*

A function $C\left(P_I|P_0\right)$ satisfies Axioms 1-4 $\iff$ $C\left(P_I|P_0\right) = cKL\left(P_I; P_0\right)$ where $c > 0$ is constant and KL is the Kullback-Liebler Divergence:

$$KL\left(P_I|P_0\right) = \int P_I \log \frac{P_I}{P_0}$$

We will first prove the easier part of the Theorem as a seprate Lemma

[23] Note that this includes all probabilities on countable sets, and continuous densities with bounded and well-defined derivative which includes the normal distribution, exponential, uniform, and many more.

Lemma 1. *cKL satisfies A1-A4.*

Proof. A1: Need to prove $\int \pi_I^n \log \frac{\pi_I^n}{\pi_0^n} - \pi_I \log \frac{\pi_I}{\pi_0} \to 0$. Rewriting as

$$\int \pi_I^n \log \frac{\pi_I^n}{\pi_I} - \pi_I^n \log \frac{\pi_0^n}{\pi_0} + (\pi_I^n - \pi_I) \log \frac{\pi_I}{\pi_0}$$

The first two components go to zero since $\frac{\pi_i^n}{\pi_i} \to 1$ uniformly. The third component goes to zero since $\log \frac{\pi_I}{\pi_0}$ is bounded and $(\pi_I^n - \pi_I) \to 0$

A2: Gibbs inequality

A3: Using integration by substitution.

A4: Writing the PDF $\pi(x)$ as $\pi(x_1)\pi(x_2|x_1)$

$$\begin{aligned} KL(P_I; P_0) &= \int_{X_1}\int_{X_2} \pi_I(x_1)\pi_I(x_2|x_1) \log \tfrac{\pi_I(x_1)\pi_I(x_2|x_1)}{\pi_0(x_1)\pi_0(x_2|x_1)} \\ &= KL(P_I(x_1); P_0(x_1)) + \int_{X_1} \pi_I(x_1) KL(P_I(x_2|x_1); P_0(x_2|x_1)) \\ &= KL(P_I(x_1); P_0(x_1)) + E_{X_1}[KL(P_I(x_2|x_1); P_0(x_2|x_1))] \end{aligned}$$

□

Next, we will use the following lemma from Hobson (1969):

Lemma 2. *Assume X is finite such that $P_0 = (p_1, ..., p_n)$ and $P_I = (q_1, ..., q_n)$*

$C(P_I|P_0)$ satisfies the following five conditions:

H1: $C(P_0|P_0) = 0$

H2: C is continuations in every $p_1, ..., p_n, q_1, ..., q_n$

H3: If $P_0 = \left(\frac{1}{n}, ..., \frac{1}{n}\right)$, $P_I = \left(\frac{1}{k}, ..., \frac{1}{k}, 0, ..., 0\right)$ with $k < n$ then $C(P_I|P_0) > 0$

H4: Let $\varphi_n \in S_n$ be a permutation. Then $C((q_1, ..., q_n) \mid (p_1, ..., p_n)) = C\left(q_{\varphi(1)}, ..., q_{\varphi(n)} | p_{\varphi(1)}, ..., p_{\varphi(n)}\right)$

H5: Mark by $\pi_1 = \sum_{i=1}^k p_i$, $\pi_2 = \sum_{i=k+1}^n p_i$,$\vartheta_1 = \sum_{i=1}^k q_i$,$\vartheta_2 = \sum_{i=k+1}^n q_i$. for $k < n$. Then
$C(P_I|P_0) = C((\vartheta_1, \vartheta_2) \mid (\pi_1, \pi_2)) + \vartheta_1 C\left(\left(\frac{q_1}{\vartheta_1}, ..., \frac{q_k}{\vartheta_1}\right) \mid \left(\frac{p_1}{\pi_1}, ..., \frac{p_k}{\pi_1}\right)\right) + \vartheta_2 C\left(\left(\frac{q_{k+1}}{\vartheta_2}, ..., \frac{q_n}{\vartheta_2}\right) \mid \left(\frac{p_{k+1}}{\pi_2}, ..., \frac{p_n}{\pi_2}\right)\right)$
$\iff$ C is KL-divergence (times a constant)

Proof. Hobson (1969) □

We will first prove Theorem 1 for the case where X is countable, by showing our axioms are equivalent to the Hobson conditions.

Lemma 3. *Let P_0, P_I be probability distributions on a countable space X. Then a function $C(P_I|P_0)$ satisfies Axioms 1-4 $\iff$ $C(P_I|P_0) = cKL(P_I; P_0)$*

Proof. $\Rightarrow$:

We will show Axioms A1-A4 $\Rightarrow$ H1-H5 and so C is KL-divergence.

H1, H3 from A2.

H2 from A1.

H4 from A3.

H5 from A4 with X_1 a dummy variable for $1, .., k$ or $k+1, ..., n$.

$\Leftarrow$

From Lemma 0 $\square$

We can now prove Theorem 1.

Proof. If X is countable, by Lemma 2.

Else, for each n, divide Xspace into a set of mutually exclusive events $\{A_i\}_{i=1}^{\infty}$ such that for any two points $x^0, x^1 \in A_i$, $dist\left(x^0, x^1\right) \leq \frac{K}{n}$ for some constant K and $\bigcup A_i = X$.

Define P_I^n to be a distribution on X such that $P_I^n\left(A_i\right) = P_I\left(A_i\right)$ and the conditional distribution $P_I^n\left(x|A_i\right) = P_0\left(x|A_i\right)$.

Define X_1^n a random variable over X with $X_1^n\left(x\right) = i$ if $x \in A_i$. Define X_2^n such that $X = X_1^n \times X_2^n$ $X_1^n \perp X_2^n$. From Additivity

$$C\left(P_I^n|P_0\right) = C\left(P_I^n\left(X_1^n\right)|P_0\left(X_1^n\right)\right) + E\left[C\left(P_I^n\left(X_2^n|X_1^n\right)|P_0\left(X_2^n|X_1^n\right)\right)\right]$$

From A2 $C\left(P_I^n\left(X_2^n|X_1^n\right)|P_0\left(X_2^n|X_1^n\right)\right) = 0$. Since X_1 is defined over a countable set, $C\left(P_I^n\left(X_1^n\right)|P_0\left(X_1^n\right)\right) = KL\left(P_I^n\left(X_1^n\right)|P_0\left(X_1^n\right)\right)$ so together

$$C\left(P_I^n|P_0\right) = KL\left(P_I^n\left(X_1^n\right)|P_0\left(X_1^n\right)\right)$$

Next, we'll show that the densities ratio satisfies $\left|\frac{\pi_I^n}{\pi_I}\right| \to 1$ uniformly. Note that $\left|\frac{\pi_I^n}{\pi_I}\right| = \left|\frac{\pi_I^n\left(x_2^n|x_1^n\right)}{\pi_I\left(x_2^n|x_1^n\right)}\right|$. Since π_I is Lipschitz continuous, if $\xi_1, \xi_2 \in A_i$ then $\left|\pi_I\left(\xi_1\right) - \pi_I\left(\xi_2\right)\right| \leq \frac{K}{n}$ for some $K < \infty$.

Note also that $\pi_I^n\left(x\right) = \pi_0\left(x\right)\frac{P_I(X_1(x))}{P_0(X_1(x))}$. Since π_0 is Lipschitz continuous and $\frac{P_I(X_1(x))}{P_0(X_1(x))}$ is bounded (by the regularity conditions), π_I^n is also Lipschitz and so the same argument applies. Hence the distributions $\pi_I^n\left(x_2^n|x_1^n\right), \pi_I\left(x_2^n|x_1^n\right)$ both uniformly converge to a uniform density that equals $\left|A_i\right|^{-1}$ so the ratio goes uniformly to 1.

Hence by A1 $C\left(P_I|P_0\right) = \lim KL\left(P_I^n\left(X_1^n\right)|P_0\left(X_1^n\right)\right)$

Finally we'll show $\lim KL\left(P_I^n\left(X_1^n\right)|P_0\left(X_1^n\right)\right) = KL\left(P_I\left(X\right)|P_0\left(X\right)\right)$. By the additivity of KL

$$KL\left(P_I\left(X\right)|P_0\left(X\right)\right) - KL\left(P_I^n\left(X_1^n\right)|P_0\left(X_1^n\right)\right) = E_{X_1}\left[KL\left(P_I\left(X_2^n|X_1^n\right)|P_0\left(X_2^n|X_1^n\right)\right)\right]$$

and as shown above, due to the Lipshcitz continuity both converge uniformly to the same uniform distribution, so the KL-divergence is 0.

The other direction was proven in Lemma 0. □

A.3 Proof of Proposition 2

Proof. Write

$$\mu_1 = \mu_0 + \nu$$

For simplicity assume that $\mu_0 = 0$, hence X is demeaned. We assume

$$f(X) = X\beta$$

And from OLS estimation

$$\widehat{\beta} = \left(X^T X\right)^{-1} X^T Y$$

Hence

$$\Omega(I) = \nu^T \left(X^T X\right)^{-1} X^T Y$$

The Kullback-Liebler divergence of two multivariate distributions is

$$D_{KL}\left(P(X|I=1)\,||P(X|I=0)\right) = \frac{1}{2}\left(\nu^T \Sigma_0^{-1} \nu + tr\left(\Sigma_0^{-1}\Sigma_1\right) + \ln|\Sigma_0| - \ln|\Sigma_1| - k\right)$$

where k is the number of variables. Taking our constants, and replacing Σ_0 with its estimator $\left(X^T X\right)$ this is

$$\frac{1}{2}\left(\nu^T \left(X^T X\right)^{-1} \nu + tr\left(\left(X^T X\right)^{-1}\Sigma_1\right) - \ln|\Sigma_1|\right)$$

Together, we maximize

$$\max_{\nu,\Sigma_1} \nu^T \left(X^T X\right)^{-1} X^T Y - \frac{\lambda}{2}\left(\nu^T \left(X^T X\right)^{-1} \nu + tr\left(\left(X^T X\right)^{-1}\Sigma_1\right) - \ln|\Sigma_1|\right)$$

The first thing to notice is that ν and Σ_1 are not interacting in this expression. Therefore this can be written as two separate exercises.

The solution $\Sigma_1 = \Sigma_0$ is from the fact that D_{KL} is minimized when the distributions are equal, and Σ_1 doesn't effect $\Omega(I)$.

To get ν we solve

$$\max_{\nu} \nu^T \left(X^T X\right)^{-1} X^T Y - \frac{\lambda}{2}\left(\nu^T \left(X^T X\right)^{-1} \nu\right)$$

FOC is

$$\left(X^T X\right)^{-1} X^T Y - \lambda \left(X^T X\right)^{-1} \nu = 0$$

$$\nu = \frac{1}{\lambda} X^T Y = COV\left(X, Y\right)$$

□

A.4 Proof of Proposition 3-5:

We will prove Proposition 5 and Proposition 3 and 4 would be specific cases and so immediately follow. We will add weights and controls.

:Assume some $\pi_0(x)$ with finite support, $f(x_i) = y_i$ and Assumptions 1 A1-A4 hold. Assume also that a vector of moments of X marked by $m(X)$ is fixed such that $E[m(X)|P_0] = E[m(x)|P_I]$ than

$$\pi_I(x_i) \propto \pi_0(x_i) \exp\left(\frac{1}{\lambda} y_i\right) \prod_j \exp\left(\frac{\gamma_j}{\lambda} m_j(x_i)\right)$$

Proof: The maximization problem in Lagarangian form is

$$\max_{\pi_I} \sum_i \left(\pi_I(x_i) y_i - \lambda \pi_I(x_i) \frac{\log \pi_I(x_i)}{\log \pi_0(x_i)} + \sum_j \gamma_j m_j\left(\pi_I(x_i) - \pi_0(x_i)\right) + \delta\left(\pi_I(x_i) - \frac{1}{N}\right)\right)$$

FOC are

$$y_i - \lambda \frac{\log \pi_I(x_i)}{\log \pi_0(x_i)} - \lambda + \sum \gamma_j m_j(x_i) + \delta = 0$$

and with simple algebra we reach the above expression. QED.

Proposition 3 & 4 follow directly from this Lemma. Note that this Lemma also is able to accomodate weights. For instance if weights are marked as $\omega(x_i)$ then the empirical distiribution is simply $\pi_0(x_i) = \omega(x_i)$ and some the value of π_I is proportional to $\omega(x_i) \exp\left(\frac{1}{\lambda} y_i\right) \prod_j \exp\left(\frac{\gamma_j}{\lambda} m_j(x_i)\right)$. Similarly weights can be added to proposition 4.

Note that the values of γ_i will be selected such that $E[m(X)|P_0] = E[m(x)|P_I]$.

Proof Theorem 6.2: Since the cost function satisfies axioms 1-3, it is an f-divergence

(Csiszer, 1963). Hence we can write the maximiation problem as

$$\max_{\pi_I(x_i)} \int \pi_I(x_i)\, y(x_i) - \lambda \pi_0(x_i)\, f\left(\frac{\pi_I(x_i)}{\pi_0(x_i)}\right) - \delta\left(\pi(x_i)_I - 1\right) dx_i$$

and taking FOC we get that for every x_i with $\pi_I(x_i) > 0$ (and also $\pi_I(x_i) < 1$ if P_I is discrete)

$$f'\left(\frac{\pi_I(x_i)}{\pi_0(x_i)}\right) = \delta + y_i$$

and since f' is monotonically increasing

$$\frac{\pi_I(x_i)}{\pi_0(x_i)} > \frac{\pi_I(x_j)}{\pi_0(x_j)} \iff y_i > y_j.$$

QED.

A.5 Proof of Theorem 2

Theorem. *Define I_D as the optimal intervention that targets only X_D within the budget constraint, holding the distribution of X_{-D} fixed. Under Assumptions 1, Axioms 1-4 a constant λ and $|Y|$ bounded*

$$\lim_{\delta \xrightarrow{p} 0} |\pi_{I_P}(x) - \pi_0(x)| = 0$$

and

$$\lim_{\delta \xrightarrow{p} 0} \Omega(Y) = 0$$

Proof. The maximization problem is the same as previously. The solution, conditional on keeping $\pi_0(X \backslash X_D)$ unchanged is

$$\pi_{I_D}(x_D|x_{-D}) = \frac{\pi_0(x_D|x_{-D}) \exp\left(\frac{1}{\lambda} y(x_D)\right)}{\int \pi_0(x_D|x_{-D}) \exp\left(\frac{1}{\lambda} y(x_D)\right)}$$

$\int \pi_0(x_D|x_{-D}) \exp\left(\frac{1}{\lambda} y(x_D)\right) \overset{\delta \to 0}{\to} E\left[\exp \frac{1}{\lambda} Y | \varphi(x_{-D})\right]$ by Portmanteau lemma. $\exp\left(\frac{1}{\lambda} y(x_p)\right)$ is bounded by assumption. Hence, since for all $x_D \neq \varphi(x_{-D}) \to 0$, it implies that $\pi^n_{I_D}(x_D|x_{-D}) \to 0$. Hence $\pi^n_{I_D}(\varphi(x_D)|x_{-D})$ converges to 1 for discrete distributions and ∞ for continuous distribtuions.

Since the distribution π_0 doesn't change $\Omega(Y) = |E[Y|P_{I_D}] - E[Y|P_0]| \to 0$ □

B Gerenraliziation of Axioms

B.1 Beyond IR

Say we know the cost of some subest of X variables which we note as X_1. In that case we can replace the invariance of representation with the following axiom.

Define **Full knowledge of cost for** X_1 as

1. Every intervention I can be written as $I = I_1 + I_2$ with $I_1 \in \mathcal{I}(X_1), I_2 \in \mathcal{I}^{\perp}(X_1)$ (see definition in axiom 4)
2. $\forall I \in \mathcal{I}(X_1)$: $C(P_I|P_0) = \chi(P_I)$ for a known function χ
3. $\forall I_1 \in \mathcal{I}(X_1), I_2 \in \mathcal{I}^{\perp}(X_1)$ - $C(P_{I_1+I_2}|P_{I_1})$ satisfies invariance of representation.

Examples: This axiom is satisfied if there is a subest of variables that cannot be changed, in which case $\chi(X_1) = \infty$. We will use the for variables such as race, and gender which will effectively only use them as controls.

Note that this axiom has several important restrictions. If X_1 variables are affecting other variables, this will be a violation of the first requirement. If there is an intervention that affects both X_1 and $X_1^{\perp}$, this would also be a violation. A violation of two would occur if, for instance, there exist other ways to intervene in X_1 which we don't know how to price.

Theorem: Under axioms 1,2,3*,4 the minimum cost of any intervention I is

$$C(P_I|P_0) = \chi(P_{I_1}) + D_{KL}(P_I|P_{I_1})$$

where $I_1 \in \mathcal{I}(X_1), I_2 \in \mathcal{I}^{\perp}(X_1)$ and $I = I_1 + I_2$

Proof: Based on 1, we know that $\exists! I_1, I_2$ s.t. $I = I_1 + I_2$. From additivity, we know that $C(P_I|P_0) = C(P_I|P_{I_1}) + C(P_{I_1}|P_0)$. From full knowledge of X_1, $C(P_I|P_0) = \chi(P_I)$. Since $C(P_I|P_I)$ satisfies all orginal axioms, it must be that $C(P_I|P_{I_1}) = D_{KL}(P_I|P_{I_1})$.

B.2 Beyond Additivity - Examples

One well-known divergence is total variation, where $f(u) = \frac{1}{2}|t-1|$

$$D_{TV}(P_I|P_0) = \frac{1}{2}\int |\pi_I(x) - \pi_0(x)|$$

This particular f-divergence is also a measure of distance as it satisfies symmetry and the triangle inequality. Using this with the non parametric estimation of P_0 and f we get the following maximiation problem

$$\max_{p(x_i)} \sum p(x_i) \log w_i - \lambda \sum \left| p(x_i) - \frac{1}{N} \right|$$

with $\sum p(x_i) = 1$. The solution for this problem, increases $p(x_i)$ only for $i = \arg\max w_i$ and gradually decreases the probability for the lowest income person with positive probability, until it hits zero. To see this, note that any other distribution will not be optimal. If $p(x_j) > \frac{1}{N}$ for another x_j (where w_j is the not the maximum), a small shift of probabailty to the x_i on the expense of x_j will increase outcome for the same cost. Similarly, if $p(x_j) < \frac{1}{N}$ where $\exists x_k$ such that $p(x_k) > 0$ and $w_j > w_k$, a small shift in probability from x_k to x_j will increase outcome for the same cost.

Therefore, we also get a similar reweighting procedure, but a more denegenrated one.

Another f-diveregence that is commonly used, though probably less that TV and KL is Pearson χ^2 divergence:

$$D_P = \int \frac{\left(p(x_i) - \frac{1}{N}\right)^2}{p(x_i)}$$

This is somewhat similar to an L_2 norm, but normalized to maintain the invariance to representation, so it is not a measure of distance. The FOC for this are

$$\log w_i - \lambda \frac{p(x_i)^2 - \frac{1}{N^2}}{p(x_i)^2} - \delta = 0$$

for every $1 > p(x_i) > 0$, and hence

$$p(x_i) = \frac{1}{N} \left(const - \frac{\log w}{\lambda} \right)^{-\frac{1}{2}}$$

Once again we get a reweighting exercise, this time with more continuous weights, where λ sets how much weigh we would put on high versus low income.

Overall, we get that different choices would lead to qualitatively similar methods, of reweighting that is based

C Simulations Details

We run three types of simulations, each type producing 1000 observations (similar to what we have in the data).

C.1 Data Generating Process

We use three simulations in this paper, which we will now describe.

C.1.1 Main Simulation

Our first simulation, produces Figure 9a. This simulation includes 10 indirect variables X_I and 10 direct variables X_D. The final outcome Y is a function of the direct variables

$$Y = f(X_D) + \varepsilon$$

The direct variables are fully set by the indirect

$$X_D^i = g_i(X_I) \tag{4}$$

The indirect variables are drawn from a normal distribution.

$$X_I^i \sim N(0, 1)$$

so the simulation can be described in the following diagram:

$$\begin{array}{ccccc} X_I & \xrightarrow{g} & X_D & \xrightarrow{f} & Y \\ & & & & \uparrow \\ & & & & \varepsilon \end{array}$$

Specifically, g_i are

$$g_i(X_I) = \left(\sum_j \alpha_j \left(1 + \left|X_I^j\right|\right)^\rho \right)^{\frac{1}{\rho}}$$

where α_j and ρ are drawn randomly at each simulations such that $\rho \leq 1$ and $\sum \alpha_j = 1$. This is similar to a CES function that can be both complimentary or substitutional. This structure generates a non-monotonic relationship between the X_I variables and the X_D variables, and through them to the outcome variables Y.[24] We choose α_j such that $\alpha_j = 0$

[24] We add 1 to $\left|X_I^i\right|$ to avoid values close to zero where $|\nabla g| \to \infty$

for exactly seven randomly chosen variables and $\alpha_j > 0$ for the remaining three variables. Hence, each X_D is a function of only three specific $X_I$variables. This feature generates a more skewed distribution of the impact of each X_I variable.

We construct f as a nested CES function which allows for substitutability and complementarity between X_D variables at the same simulation.

$$f\left(X_D\right) = \log\left(\sum_{k=1}^{2}\beta_k\left(\sum_{l=1}^{5}\alpha_l\left(1+X_D^i\right)^{\nu_k}\right)^{\frac{\rho}{\nu_k}}\right)^{\frac{1}{\rho}}$$

For the parameters in f, g we define $\rho = 1 - Z$ where $Z \sim Exp\left(\log 2\right)$ such that ρ is between $(-\infty, 1]$ with $P\left(\rho > 0\right) = P\left(\rho < 0\right) = \frac{1}{2}$ and similarly for ν. We draw α and β by drawing $n-1$ numbers from the $[0,1]$ interval using a Uniform distribution. We then use these variables to divide the $[0,1]$ interval randomly into n intervals, and define α_i (β_i) as the length for the ith interval.

Finally, ε is drawn randomly from a distribution

$$\varepsilon \sim N\left(0, \sigma_f\right)$$

where σ_f is the standard deviation of $f\left(X_D\right)$.

To add some real-life complication, we add some noise variables that are unrelated to Y. We will simulate 10 additional noise variables X_N with

$$X_N^i \sim N\left(0,1\right)$$

that are not affecting Y in any way. These variables don't have any information on Y so with infinite data all methods would mark them as unimportant. But with finite data, some methods might mistakenly classify them as important variables, as we'll see. The full data set includes all 30 X variables $X = \left(X_D, X_I, X_N\right)$.

C.1.2 Including Side-Effects

In the second simulation (figure 9b) we add additional set of 10 variables X_S such that

$$X_S^i = h_i\left(X_I\right)$$

where h_i has the same functional form as g_i. The causal structure can now be described with the following diagram

$$\begin{array}{ccccc} X_I & \xrightarrow{g} & X_D & \xrightarrow{f} & Y \\ \downarrow h & & & & \uparrow \\ X_S & & & & \varepsilon \end{array}$$

In this setting we have 40 total X variables $X = (X_D, X_I, X_N, X_S)$

C.1.3 Direct Malleable Effects

The third and final simulation (figure 9c) generates two distinct sets of X_D variables. X_{D_1} which are identical to our previous X_D variables and are generated by the X_I variables using the functions g_i, and X_{D_2} that are direct variables that can be changed directly in an intervention and are drawn from a standardize Normal distribution

$$X_{D_2} \sim N(0, 1)$$

The final outcome is now

$$Y = \log\left[(1-\alpha) f_1(X_{D_1}) + \alpha f_2(X_{D_2})\right] + \varepsilon$$

where f_1 and f_2 are nested logits like before. In this simulation there are no additional noise variables to the 40 X variables are $X = (X_{D_1}, X_{D_2}, X_I, X_S)$.

C.2 Simulating an Intervention

Each intervention I chooses a subset s_I of variables that are the top variables from each method. We define "Score", the effect of an intervention as the marginal gain from changing these variables, under the assumption that X_D cannot be changed and that all X_I are changed an equal and infinitesimal amount, at the optimal direction:

$$Score(I) = E\left[\left|\frac{dY}{dI}\right|\right] = E\left[\left|\frac{dY}{dX_D}\frac{dX_D}{dX_I}\frac{dX_I}{dI}\right|\right] = \sum_{X_I^i \in s_I} E\left[\left|\frac{dY}{dX_D}\frac{dX_D}{dX_I}\right|\right]$$

For every value of $m = 1, ..., 30$ and for every simulation we find the subset s_I of variables that will create the largest impact. Mark this intervention by I_m^*. Note that for $m \leq 10$ these will always be a subset of X_I variables, and for $m > 10$ there is no difference in score ($Score(I_m^*) = Score\left(I_{m+1}^*\right)$ for $m \geq n_1$).

At each simulation we ran our non parametric method, random forests, OLS and raw correlations. Let $Score(I_m^{NP/RF/OLS/Cor})$ be the effect of an intervention on the top m variables based on each of the method we are measuring.

Figure 9-11 shows the effect of each intervention that is based on the top m variables from each method, divided by the effect of the optimal intervention

$$\frac{Score(I_m^{NP/RF/OLS/Cor})}{Score(I_m^*)}$$

D Data Appendix

D.1 Pilot Survey Implementation Guide

D.1.1 Project Timeline

March 28, 2017 - April 30, 2017	Phone Survey
April 6, 2017 - August 2, 2017	Address Based Survey

D.1.2 Sample Design

Population The target population for the study consisted of non-institutionalized persons age 18 and over living in Los Angeles County, California (hereafter: L.A. County).

Sampling Frame Numbers for the landline sample were drawn by Survey Sampling International with equal probabilities from active blocks (area code + exchange + two-digit block number) that contained one or more residential directory listings. The cellular sample was drawn by Survey Sampling International through a systematic sampling from 1000 blocks dedicated to cellular service according to the Telcordia database. The sampling frame excluded non-telephone households. The address-based sample was drawn by Marketing Systems Group using a stratified cluster design. Census tracts were used as primary sampling units (PSU) with a systematic probability proportionate to size selection. Within each of the 110 sampled PSUs, 55 addresses were selected for a total of 5,500 addresses from which 3 self-representative replicates were created. The sampling frame excluded vacant, seasonal, educational, P.O. boxes not flagged as the only address where the owner receives mail, and drops (multi-residence dwellings with no unit number). The strata are described below.

Stratification

The ABS sample stratified the universe of Census tracts in L.A. County by poverty rates. We obtained tract-level poverty estimates from the American Community Survey 2015 5-year Summary File, using 100% of the Federal poverty line as our benchmark. County-level poverty rates were then Z-scored. Tracts with poverty rates more than one standard deviation below the mean were assigned to the Low Poverty (hereafter, "Low") stratum, tracts with poverty rates more than one standard deviation above the mean were assigned to the High Poverty (hereafter, "High") stratum, and tracts with poverty rates between -1 and 1 standard deviations were assigned to the Medium Poverty ("Medium") stratum. Tracts for which no information on poverty is available (those with very few residents) were assigned to the Medium stratum.

Respondent Selection For the telephone and CAPI modes, interviewers asked to speak with an adult age 18 or older, living in the household. The web invitations and paper surveys were sent via USPS with the assumption that adults were opening and reading the households mail.
Screening and Eligibility The survey screened for respondents age 18 or older who live in L.A. County. Anyone providing a ZIP Code outside of L.A. County was screened out and were not eligible for the survey.

D.1.3 Comparison Between Address Based Sampling and Phone Based Sampling

	Address-Based	Phone-Based	p-value
	(1)	(2)	(1) = (2)
Age	47.406	49.157	0.220
Female	0.606	0.557	0.202
Race			
White	0.425	0.444	0.633
Black	0.094	0.129	0.160
Hispanic	0.358	0.373	0.694
Asian	0.141	0.064	0.002
Other race	0.059	0.041	0.302
Education			
Less than HS	0.050	0.064	0.453
Some HS, Incomplete	0.056	0.074	0.357
High School/GED/Incomplete College	0.294	0.332	0.301
Two Year Associate Degree	0.094	0.077	0.448
Bachelor or Graduate Degree	0.506	0.453	0.183
Household Income			
Less then 15,000	0.146	0.189	0.157
15,000 to less than 25,000	0.109	0.148	0.164

	Address-Based	Phone-Based	p-value
25,000 to less than 35,000	0.097	0.110	0.617
35,000 to less than 50,000	0.128	0.076	0.040
50,000 to less than 75,000	0.152	0.121	0.282
75,000 to less than 100,000	0.125	0.095	0.250
100,000 or more	0.243	0.261	0.612
Observations	300	343	

D.2 Survey Implementation Guide

D.2.1 Project Timeline

September 5, 2017 - April 9, 2018	Phone screens conducted
October 30, 2017	Mailed out study brochures
November 20, 2017 - May 7, 2018	In-person interviews conducted
May 21, 2018	Final dataset received

D.2.2 Sample Design

D.2.3 Population

The target population for the study consisted of adults age 18 and over who self-reported having grown up poor and currently living in Shelby County, TN, Tulsa County, OK, or Jefferson and Orleans Parishes, LA.

D.2.4 Sample

Sampling Frame Numbers for the landline sample were drawn by Survey Sampling International (SSI) with equal probabilities from active blocks (area code + exchange + two-digit block number) that contained one or more residential directory listings. The cellular sample was drawn by Survey Sampling International through a systematic sampling from 1000 blocks dedicated to cellular service according to Telcordia, an FCC approved national telephone database administrator. The sampling frame excluded non-telephone households. Each sampling frame is described in detail below.

Landline sample When creating its landline RDD database, SSI starts with a computer file of over 53 million directory-listed households. Using area code and exchange data regularly obtained from Telcordia and additional databases, this file is subjected to an extensive cleaning and validation process to ensure that all exchanges are currently valid, assigned to the correct area code, and fall within an appropriate set of ZIP Codes. Telephone exchanges and 100 blocks (i.e., the last two digits of the telephone number) that contain one or more listed residential telephone numbers are considered valid and are represented in the database. The RDD database is formed of all telephone numbers having such valid exchanges and working blocks.

Each exchange is assigned to a single county. For those exchanges that cross county and/or state lines, the exchanges are assigned in their entirety to the county with the

highest number of listed phones within that exchange. Abt Associates selects random digit samples from the RDD database using the Random A procedure. Random A is an SSI term denoting a systematic sample of random digit telephone numbers selected with equal probability across all working blocks. Within a county, the sampling interval is calculated by dividing the number of working blocks by the number of sampling points requested. Abt Associates uses a working block value of one to minimize under-coverage.

Cell Phone Sample Mobile samples are selected from a database that contains all possible numbers in 100-blocks dedicated to wireless service and 100-blocks providing shared services but that have no directory-listed telephone numbers. SSI selects EPSEM (equal probability of selection methodology) samples selected from the cell phone number database. Blocks are in ascending order by exchange and block number within exchange, within county. Once the quota has been allotted to all the counties in the frame, a sampling interval is calculated for each county by summing all the working blocks in the county and dividing that sum by the number of sampling points assigned to the county. From a random start between zero and the sampling interval, blocks are systematically selected from each county. Once a block has been selected, a two-digit random number in the range 00-99 is appended to the block to form a ten digit telephone number.

In order to more efficiently reach cell phone respondents, the cell sample was appended with activity code information by SSI. These activity codes provided information about the working status of each number in the cell phone sample. In the cell sample, 67% of numbers were flagged as "active" and 33% were flagged as "inactive". Completely excluding all cell phone numbers classified as "inactive" from dialing could result in some coverage bias if some of those numbers were actually active, as observed in the pilot study. For this reason, cell phone numbers flagged as "inactive" were sub-sampled at approximately 50% rate.

Billing ZIP Code-Matched Cell RDD Sample Using billing ZIP Code-matched cell RDD sample can help to address the low geographic eligibility rate observed in the pilot, reducing screening costs. ZIP-matching is, however, only possible for cell phones with billing records. (The precise details of ZIP-matching methodology are proprietary and held closely by SSI and its competitors.) Therefore, SSI appended billing ZIP Code to every cell phone selected in the sample. From that we observed that 30.6% of the numbers were matched to a billing ZIP code inside the target areas, 11.8% were matched to a

billing ZIP code outside the target areas and 57.6% of the numbers were not matched to a billing ZIP code. In order to improve the efficiency of the sample, all phone numbers matched to a billing ZIP code outside the targeted areas were excluded from the sample and phones numbers that were not matched to any billing ZIP code were sub-sampled at approximately 65%. All phone numbers matched to a billing ZIP code inside the targeted areas were eligible to be dialed.

Respondent Selection For the landline sample, interviewers asked to speak with an adult age 18 or older, living in the household. For the cell sample, interviews were conducted with the person who answered the phone and was age 18 or older.

Screening and Eligibility The telephone screener first confirmed the age of the respondent and the zip code where the individual lived. If the respondent was 18 or over and lived in one of the eligible geographic areas, the survey then also screened for respondents who self-identified as having "grown up poor". Persons who reported not having grown up poor were not eligible for the in-person extended interview, but their basic demographic information, such as gender, age, education and race/ethnicity, were collected for weighting purposes.

D.2.5 Number of Respondents by Market

	Shelby	Tulsa	Jefferson Orleans	Total
	(1)	(2)	(3)	(4)
Total Phone Screens	2281	2237	1941	6459
Ineligible from Phone Screens	1334	1336	1172	3842
Eligible but Declined	526	489	375	1390
Eligible and Agreed	421	412	394	1227
In-Person Interviews	308	308	312	928
Removed Well off Respondents	300	301	299	900

D.3 Data Description and Coding of Variables

D.3.1 Survey Variables and Indices

Outcome Variables:

1. Income:

   (a) Household Income – This outcome was collected from the survey question, “What was your total annual household income last year, in 2016, before taxes?" When respondent received benefits from the government, responses were collected from the question, “What was your total annual household income last year, in 2016, before taxes, as you reported to the government?" If the respondent did not provide an answer to the above two questions, they were asked to give the range in which their household income existed

   - Under $10, 000
   - $10,000 to $20,000
   - $20,000 to $30, 000
   - $30,000 to $40,000
   - $40,000 to $50,000
   - $50,000 to $75,000
   - $75,000 to $100,000
   - $100,000 to $150,000
   - More than $150,000

   If the range was specified, then household income for different ranges were taken as

   - $10,000
   - $15,000
   - $25,000
   - $35,000
   - $45,000
   - $62,500
   - $87,500
   - $125,000
   - $150,000

(b) Adjusted Household Income – Adjusted household income was calculated using the formula used for the Census Supplement of Poverty Measure.

i. For a household without children

$$N = (adults)^{.5}$$

ii. Single parent household

$$N = (1 + 0.8 * \text{first child} + 0.5 * \text{other children})^{.7}$$

iii. All other

$$N = (adults + 0.5 * children)^{.7}$$

Then adjusted income is

$$\frac{HH - Income}{N}$$

(c) Individual Income – This outcome was collected from the survey question, "If you are not the only income earner in your household, what was your individual annual income last year, in 2016, before taxes?" When respondent received benefits from the government, responses were collected from the question, "If you are not the only income earner in your household, what was your individual annual income last year, in 2016, before taxes, as you reported to the government?" If the respondent did not provide an answer to the above two questions, they were asked to give the range in which their individual income existed and were recoded in the same way as household income above.

2. Adult Physical Health: This outcome was coded from the question, "In general, how is your health: excellent, very good, good, fair or poor?" where 1 = poor, 2 = fair, 3 = good, 4 = very good, and 5 = excellent.

3. Adult Drug and Alcohol Use: This outcome was calculated as the total of 4 questions –

- Have you ever felt that you ought to cut down on your drinking or drug use?
- Have people annoyed you by criticizing your drinking or drug use?
- Have you ever felt bad or guilty about your drinking or drug use?
- Have you ever had a drink or used drugs first thing in the morning to steady your nerves or get rid of a hangover?

We coded each question to be 1 if they responded “yes" and if their earliest episode happened after the age of 18; or if they responded “yes" and their most recent episode happened after the age of 18. Variables were coded as 0 if they responded “no", or if they responded “yes" but their earliest and most recent episode happened before the age of 18 years.

4. Adult Mental Illness: This outcome is calculated as the total of standardized responses to the following question –

   For each of these, please indicate how often you have been bothered by it over the last 2 weeks.

   - Feeling nervous, anxious or on edge
   - Not being able to stop or control worrying
   - Trouble relaxing
   - Being so restless that it is hard to sit still
   - Becoming easily annoyed or irritable
   - Feeling afraid as if something awful might happen
   - Little interest or pleasure in doing things
   - Feeling down, depressed or hopeless
   - Trouble falling asleep, staying asleep or sleeping too much
   - Feeling tired or having little energy
   - Poor appetite or overeating
   - Feeling bad about yourself or that you are a failure or have let yourself or your family down
   - Trouble concentrating on things, such as reading the newspaper or watching television
   - Moving or speaking so slowly that other people could have noticed. Or, the opposite being so fidgety or restless that you have been moving around a lot more than usual
   - Thoughts that you would be better off dead or of hurting yourself in some way.

   where each statement’s response is coded as 1 = not at all, 2 = several days, 3 = more than half of days, 4 = nearly every day.

Demographics:

1. Gender - Coded from phone screener as “male" or “female".
2. Race - Coded from phone screener as “black", “hispanic", “white" or “other".
3. Education - Coded from phone screener as
    - No high school
    - Incomplete high school
    - High School/GED/Incomplete College
    - Two year Associate Degree
    - Bachelor or some Post-Graduate Degree
4. Parental Income -
    - Would you say your family during that time was pretty well off financially, about average, or poor?
    - While you were growing up, from birth until age 16, did financial difficulties ever cause you or your family to move to a different place?
    - From birth until age 16, was there a time when you or your family received help from relatives because of financial difficulties?
    - From birth until age 16, was there a time of several months or more when your father had no job?
    - Did your mother ever get Aid to Families with Dependent Children or welfare from your birth until age 16?
    - How often do you remember the following happening to your parents or guardian until you were age 16? Would you say always, often, sometimes, rarely, or never?
        - Being unable to find child care or being forced to take a child out of child care because they couldn’t pay?
        - Falling behind in rent or mortgage payments?
        - Falling behind in gas, electric, or phone bills?
        - Being unable to pay for adequate transportation to get to work or school?
        - Being unable to get medical care because of the cost?
        - Having trouble paying a credit card balance?

- Having too little money to buy enough food?
- Being a victim of a crime?
- Having a problem with alcohol or drug abuse?

Childhood Experience:

1. Lifetime Trauma Before 18

   - Before you were 18 years old, did you ever have to do a year of school over again?
   - Before you were 18 years old, were you ever in trouble with the police?
   - Before you were 18 years old, did either of your parents drink or use drugs so often that it caused problems in the family?
   - Before you were 18 years old, were you ever physically abused by either of your parents?

2. Any Lifetime Trauma - Please indicate whether the event occurred at any point in your life

   - Have you ever been homeless?
   - Have you ever been in a major fire, flood, earthquake, or other natural disaster?
   - Did you ever have a life-threatening illness or accident?
   - Did your spouse or a child of yours ever have a life-threatening illness or accident?
   - Have you ever fired a weapon in combat or been fired upon in combat?
   - Has your spouse, partner, or child ever been addicted to drugs or alcohol?
   - Were you the victim of a serious physical attack or assault in your life?
   - Has a child of yours ever died?

3. Adverse Childhood Experience -

   - Did a parent or other adult in the household often swear at you, insult you, put you down, or humiliate you or act in a way that made you afraid that you might be physically hurt?
   - Did a parent or adult in the household often push, grab, slap, or throw something at you or ever hit you so hard that you had marks or were injured?

- Did an adult or person at least 5 years older than you ever touch or fondle you or have you touch their body in a sexual way or try to actually have oral, anal, or vaginal sex with you when you were a minor? (IF NEEDED: A minor is someone under the age of 18.)
- Did you often feel that no one in your family loved you or thought you were important or special or your family didn't look out for each other, feel close to each other, or support each other?
- Did you often feel that you didn't have enough to eat, had to wear dirty clothes, and had no one to protect you or your parents were too drunk or high to take care of you or take you to the doctor if you needed it?
- Were your parents ever separated or divorced?
- Was your mother or stepmother often pushed, grabbed, slapped, or had something thrown at her or sometimes or often kicked, bitten, hit with a fist, or hit with something hard or ever repeatedly hit over at least a few minutes or threatened with a gun or knife?
- Did you live with anyone who was a problem drinker or alcoholic or who used street drugs?
- Was a household member depressed or mentally ill or did a household member attempt suicide?
- Did a household member go to prison?

4. Risky Attitudes as Teenager -

    - I had trouble concentrating or paying attention
    - I lied or cheated
    - I teased others a lot
    - I disobeyed my parents
    - I had trouble sitting still
    - I had a hot temper
    - I would rather have been alone than with others
    - I hung around with kids who got into trouble
    - I disobeyed at school
    - I didn't get along with other kids

- I had trouble getting along with teachers
- First, I would like to ask you about smoking habits. As a teenager, did you smoke cigarettes?
- As a teenager, did you ever have a drink of an alcoholic beverage? By a drink we mean a can or bottle of beer, a glass of wine, a mixed drink, or a shot of liquor. Do not include childhood sips that you might have had from an older person's drink.
- As a teenager, did you ever use marijuana (that is grass or pot)?
- Excluding marijuana and alcohol, as a teenager, did you ever use any other drugs like cocaine or crack or heroin, or any other substance not prescribed for you by a doctor, in order to get high or to achieve an altered state?

5. Beliefs About Success

    - When you were young, did you believe you would grow up to be successful?
    - Was there a time in your life when help could have made all the difference?

Health Indices:

1. Mental Illness Before 16 - Before 16 did you have

    - Depression
    - Drug or alcohol problems
    - Any other emotional or psychological problems

2. Physical Illness Before 16
3. Resilience -

    - I tend to bounce back quickly after hard times
    - I have a hard time making it through stressful event
    - It does not take me long to recover from a stressful event
    - It is hard for me to snap back when something bad happens
    - I usually come through difficult times with little trouble
    - I tend to take a long time to get over set-backs in my life

4. Locus of Control

- “What happens to me is my own doing" or “Sometimes I feel that I don’t have enough control over the direction my life is taking."
- “When I make plans, I am almost certain that I can make them work" or “It is not always wise to plan too far ahead because many things turn out to be a matter of good or bad fortune anyhow."
- “In my case getting what I want has little or nothing to do with luck" or “Many times we might just as well decide what to do by flipping a coin."
- “Many times I feel that I have little influence over the things that happen to me" or “It is impossible for me to believe that chance or luck plays an important role in my life."

5. Growth Mindset

    - You have a certain amount of intelligence, and you really can’t do much to change it.
    - Your intelligence is something about you that you can’t change very much.
    - You can learn new things, but you can’t really change your basic intelligence.

6. Grit

    - New ideas and projects sometimes distract me from previous ones.
    - Setbacks don’t discourage me.
    - I have been obsessed with a certain idea or project for a short time but later lost interest.
    - I am a hard worker.
    - I often set a goal but later choose to pursue a different one.
    - I have difficulty maintaining my focus on projects that take more than a few months to complete.
    - I finish whatever I begin.
    - I am diligent.

7. Self Esteem

    - On the whole, I am satisfied with myself.
    - At times, I think I am no good at all.

- I feel that I have a number of good qualities.
- I am able to do things as well as most other people.
- I feel I do not have much to be proud of.
- I certainly feel useless at times.
- I feel that I'm a person of worth, at least on an equal plane with others.
- I wish I could have more respect for myself.
- All in all, I am inclined to feel that I am a failure.
- I take a positive attitude toward myself.

8. Self Control

    - I am good at resisting temptation.
    - I have a hard time breaking bad habits.
    - I am lazy.
    - I say inappropriate things
    - I do certain things that are bad for me, if they are fun.
    - I refuse things that are bad for me.
    - I wish I had more self-discipline.
    - People would say that I have iron self-discipline.
    - Pleasure and fun sometimes keep me from getting work done.
    - I have trouble concentrating.
    - I am able to work effectively toward long-term goals.
    - Sometimes I can't stop myself from doing something, even if I know it is wrong.
    - I often act without thinking through all the alternatives.

9. IPIP

    - I am the life of the party.
    - I feel little concern for others.
    - I am always prepared.
    - I get stressed out easily.
    - I have a rich vocabulary.

- I don't talk a lot.
- I am interested in people.
- I leave my belongings around.
- I am relaxed most of the time.
- I have difficulty understanding abstract ideas.
- I feel comfortable around people.
- I insult people.
- I pay attention to details.
- I worry about things.
- I have a vivid imagination.
- I keep in the background.
- I sympathize with others' feelings.
- I make a mess of things.
- I seldom feel blue.
- I am not interested in abstract ideas.
- I start conversations.
- I am not interested in other people's problems.
- I get chores done right away.
- I am easily disturbed.
- I have excellent ideas.
- I have little to say.
- I have a soft heart.
- I often forget to put things back in their proper place.
- I get upset easily.
- I do not have a good imagination.
- I talk to a lot of different people at parties.
- I am not really interested in others.
- I like order.
- I change my mood a lot.

- I am quick to understand things.
- I don't like to draw attention to myself.
- I take time out for others.
- I ignore my duties.
- I have frequent mood swings.
- I use difficult words.
- I don't mind being the center of attention.
- I feel others' emotions.
- I follow a schedule.
- I get irritated easily.
- I spend time reflecting on things.
- I am quiet around strangers.
- I make people feel at ease.
- I am exacting in my work.
- I often feel blue.
- I am full of ideas.

10. Diet

Family Indices:

1. Family Environment
2. Family Network
3. Relationship with Parents
4. Parenting
    - How often would your parent say that you did something that gave him/her pleasure and enjoyment?
    - How often would your parent say that you did something that greatly irritated him/her and got on his/her nerves?
    - How often did your parent read to you?

- How often did your parent physically punish you as a child, for example by a spanking?
- How often did your parent praise you as a child, by saying something like "Good for you!" "What a nice thing you did!" "Thank you!" or "That's good going!"
- How often did your parent tell you about his/her experience, by saying something like, "I saw a pretty bird outside just a little while ago", or "I exercised so hard that I got really tired", or "I was able to give some directions today to somebody that got lost", or "I really like the way the sky looks now".
- How often did you and your parent talk or play with each other, focusing attention on each other for five minutes or more, without your parent asking or telling you to do anything?
- How often did your parent tell you to do something, with an irritated or angry tone of voice?
- How often did you and your parent engage in make-believe play, where you each played the part of a character, and together made up a story to act out with each other?
- How often did you laugh with your parent?
- How often did your parent yell or speak to you in a very loud voice, with irritated or angry emotion?

5. Adults you Could Trust 100%
6. Number of Adult Relationships Trusted
7. Quality of Adult Relationships
8. Adult Living with at Age 9

Neighborhood Indices:

1. Probability of Bottom 25 in Top 20 percentile
2. Fraction with Fathers Present
3. Neighborhood Safety Index
    - How safe were the parking lots and sidewalks near your neighborhood school?

- How safe did you feel at home alone at night?
- How safe were the streets near your home during the day?
- How safe were the streets near your home at night?
- Was anyone's purse, wallet, or jewelry snatched from them?
- Was anyone threatened with a knife or gun?
- Was anyone beaten or assaulted?
- Was anyone stabbed or shot?
- Did anyone try to break into your home?

D.3.2 NLSY

The National Longitudinal Survey of Youth 1979 (NLSY79) is a nationally representative sample of 12,686 young men and women who were between 14 and 22 years old when they were first interviewed in 1979. These individuals were surveyed annually through 1994 and biannually thereafter. Covered topics include family background and demographic characteristics; household composition; schooling and aptitude information; income and assets; health conditions; alcohol and substance use; attitudes and aspirations; and more. For our analysis, we defined growing up poor to be an individual who we observe in the 2014 sample and in 1979 had a Poverty Status equal to one, an NLSY created indicator variable based on number of respondents in the household, family income and yearly poverty income guidelines established by the U.S. Department of Health and Human Services. For current income, we used the individual's total income in 2014. Our interview used the same four questions as the NLSY79 to measure Rotter's Locus of Control Scale. For additional information on the NLSY79 see the NLS Handbook. (https://www.bls.gov/nls/handbook/2005/nlshc2.pdf)

D.3.3 Dataset from Dobbie and Fryer 2013

The dataset was collected for Dobbie and Fryer, 2013 and is constructed from two sources: school-specific data collected from principal, teacher, and student surveys, lesson plans, and videotaped observations of classroom lessons; and administrative data on student demographics and outcomes from the New York City Department of Education (NYC-DOE). In 2010, the authors attempted to collect data from all charter schools in New York City with students in grades 3-8. Eligible schools were invited to participate via e-mail and phone and offered a stipend. Of the 62 eligible charter elementary schools (entry

grades of PK to fourth) and 37 eligible charter middle schools (entry grades of fifth to eighth), 26 elementary schools, and 13 middle schools chose to participate in the study. Within the set of participating schools, 19 elementary schools, and 10 middle schools were also able to provide usable admissions lottery data. A wide variety of information was collected from participating schools including details on teacher and staff development, instructional time, data driven instruction, and parent outreach obtained through a principal interview. Information on curricular rigor was coded from lesson plans collected for each testable grade level in both math and ELA. Finally, information on school culture and practices was gathered during full day visits to each school. Within each domain, an indicator variable was coded to equal to one if a school has an above median level of that input, selecting the variable or combination of variables that best captures the variation described by the qualitative literature. The administrative data from the NYCDOE included information on student race, gender, free and reduced-price lunch eligibility, behavior, attendance, and state math and ELA test scores for students in grades 3-8. Additional details are available in the Data Appendix of Dobbie and Fryer 2013.

D.4 Phone Screen Instrument

24256 Intergenerational Poverty
CATI Screener
August 24, 2017

Hello, my name is _____ and I'm calling from Abt Associates, a survey research firm on behalf of EdLabs at Harvard University (IF NEEDED: That is, the Education Innovation Laboratory). We're conducting a brief research survey about how people's background and upbringing affects their lives and future resources. It is called the "Understanding Inequality Survey." It will only take 10 minutes of your time. LANDLINE ONLY: May I please speak to an adult, age 18 or older? (IF NEW RESPONDENT, RE-INTRODUCE AND CONTINUE) CELL ONLY: Are you 18 years old or older? (if No, screen out)

Thank you, as I mentioned the survey is about your upbringing and how you are doing today. We'll start with some questions about your background and then move on to questions about your current situation.

All of your responses will be kept confidential and the reported results of the study will combine the responses of yours and others. No personal identifying information will be reported. Participation is voluntary and you can stop at any time. Some of the questions may be considered sensitive in nature and you may skip any question you prefer not to answer. This call may be monitored for quality assurance. Would you like the contact information for the researchers or the Committee on the Use of Human Subjects in Research at Harvard University, who approved this study?

[IF YES: If you have any questions about the survey, you can call 877-699-4340 or email InequalitySurvey@abtassoc.com. Additionally, this research has been reviewed by the Committee on the Use of Human Subjects in Research at Harvard University. They can be reached at 617-496-2847, 1350 Massachusetts Avenue, 9th Floor, Suite 935 Cambridge, MA 02138 and cuhs@harvard.edu.

(If cell phone: If you are now driving a car or doing any activity requiring your full attention, I need to call you back later.) If you do not have any questions, may we begin?

1. To begin, for classification purposes only, can you please tell me your age as of today? RANGE 18-98, 99=Don't know/Refused

IF Q1=99 ASK Q1a; ELSE code into appropriate category

1a. Are you between the ages of:

1. 18-24
2. 25-34
3. 35-44
4. 45-54
5. 55-64
6. 65 or more
9. (VOL) Don't know/Refused

2. To make sure all areas are represented, can you please tell me which zip code you live in?

1. [DISPLAY APPROPRIATE ZIP CODES BASED ON MARKET]
98. (VOL) Other (specify) – Screen out
99. (VOL) Don't know/Refused – Screen out

3. And how long have you lived in the [READ IN RESPONSE from Q2] zip code? Would you say… [read list]

1. Less than one year
2. 1-2 years
3. 3-4 years
4. 5-9 years
5. 10-14 years
6. 15-20 years
7. More than 20 years?
8. (VOL) My whole life
9. (VOL) Don't know/Refused

ASK IF Q3=1 or 2, ELSE SKIP TO Q4

3a. What zip code did you live in before moving to [READ IN RESPONSE from Q2]?

[Numeric entry, verify matches ZIP Code format]

IF Q3=8 OR (Q1a=1 AND Q3=7) AUTOPUNCH Q4=1 AND SKIP TO Q5; ELSE, READ:

4. Where did you spend the majority of your time living as a child, between the ages of 5 and 15? Was it… (READ LIST)

   1. The neighborhood you currently live in
   2. Another neighborhood in [Shelby County]/[Tulsa County]/[your Parish]
   3. Somewhere else in [Tennessee/Oklahoma/Louisiana]
   4. A different state
   5. Overseas
   9. (VOL) Don't know/Refused

5. How many people currently live in your household, including yourself? Please also include any children, under the age of 18.

   1. RANGE: 1-15
   99. (VOL) Don't know/Refused

ASK IF Q5>1

5a. Is there anyone living in your household who is NOT in your immediate family. If so, how many? This includes people such as extended family (cousins, grandparents), friends, and roommates.

   (IF NEEDED: By immediate family, we mean a spouse/domestic partner, any siblings, parents, or children.)

   1. RANGE 1-number from Q5
   99. (VOL) Don't know/Refused

6. How would you rate your current financial situation, would you say it's… (READ LIST)

   1. Excellent
   2. Good
   3. Only Fair, or
   4. Poor
   9. (VOL) Don't know/Refused

7. Have you, or have any immediate family members in your household ever received welfare or public assistance benefits? This includes SNAP ("snap") or food stamps, TANF ("Tan-Eff"), or other housing benefits.

   1. Yes
   2. No
   9. (VOL) Don't know/Refused

ASK IF Q7=1 (Yes)

7a. Are you or are any immediate family members in your household receiving such benefits now?

1. Yes
2. No
9. (VOL) Don't know/Refused

8. Do you consider yourself to have grown up poor?

1. Yes
2. No
9. (VOL) Don't know/Refused

9. Do you consider yourself poor now?

1. Yes
2. No
9. (VOL) Don't know/Refused

Thank you. Now I have just a few final questions to help us classify your answers.

10. I have to verify, what is your gender?

1. Male
2. Female
3. (VOL) Other
9. (VOL) Don't know/Refused

11. What is the highest level of education you have received?

1. Less than high school (Grades 1-8 or no formal schooling)
2. High school incomplete (Grades 9-11 or Grade 12 with NO diploma)
3. High school graduate (Grade 12 with diploma or GED certificate)
4. Some college, no degree (includes some community college)
5. Two year Associate's degree from a college or university or community college
6. Four year college or university degree/Bachelor's degree (e.g., BS, BA, AB)
7. Some postgraduate or professional schooling, no postgraduate degree (e.g. some graduate school)
8. Postgraduate or professional degree, including master's, doctorate, medical or law degree (e.g., MA, MS, PhD, MD, JD, graduate school)
9. (VOL) Don't know/Refused

12. What is your race? (Accept multiple responses)

1. White/Caucasian
2. Black/African American
3. Hispanic
4. Asian/Asian-American
5. Some Other Race (Specify)
6. (VOL) Native American/American Indian/Alaska Native
7. (VOL) Pacific Islander/Hawaiian
9. (VOL) Don't know/Refused

ASK Q13a OF LL SAMPLE ONLY

13a. Now thinking about your telephone use…does anyone in your household, including yourself, have a working cell phone?

1. Yes
2. No
8. (VOL) Don't know
9. (VOL) Refused

ASK Q13b OF CELL SAMPLE ONLY

13b. Now thinking about your telephone use…is there at least one telephone inside your home that is currently working and is not a cell phone?

1. Yes
2. No
8. (VOL) Don't know
9. (VOL) Refused

IF Q8>1, SKIP TO QCLASS

ASK Q14 of CELL SAMPLE ONLY

14. Those are all of the questions I have. For speaking with us today, we'd like to mail you a check for $5. All I'll need is your name and address. Please note, this will be used for your check only and will not be associated with your survey responses.

Name:
Address:
City:
State:
Zip:

15. Thank you very much for answering my questions. Based on your answers to this survey, you qualify for the next round of our study which consists of an in-person interview. If you complete the full interview, you will be paid $150 as a thank you for your time. [IF NECESSARY: The $150 will be in the form of a pre-loaded Visa or MasterCard gift card.] The survey will take about two hours to complete and many people find the questions very interesting. An interviewer will come to interview you. [IF NEEDED: The interviewer will visit your home.] Can I take your contact information so that the interviewer can contact you to make a time for the interview?

1. Agrees to participate - CONTINUE
9. Refused – SKIP TO QCLASS.

16. Great. To start, I'll just need your name and address.

[USE STANDARD ADDRESS BLOCK, FOR CELL VERSION, READ-IN INFO FROM Q14, IF REFUSED, SKIP TO QCLASS]

17. And can I please have the best phone number to reach you?

1. Enter phone [CONTINUE]
9. Refused [SKIP TO Q19]

17a. Is this a landline, cell, or work number?

1. Landline
2. Cell
3. Work

18. Is there another phone number that we can try if we can't reach you there?

1. Enter phone [CONTINUE]
9. Refused [SKIP TO Q19]

18a. Is this a landline, cell, or work number?

1. Landline
2. Cell
3. Work

IF Q17a=2 OR Q18a=2, ASK Q21, ELSE – SKIP

19. Do we have your permission to contact you via text message to your cell phone?

1. Yes
2. No

20. Can I please have your email address?

1. Provided email
2. Don't have email
9. Refused

21. Sometimes when we try to contact people, we aren't able to reach them because they may have moved or got a new telephone number. I'd like to get the contact information for 2 other people who don't live with you that would know how to reach you? Let's start with the first person.

[USE STANDARD NAME/ADDRESS TEMPLATE]

21a. What is their relationship to you?

1. Parent
2. Brother/Sister
3. Son/Daughter
4. Other Family Member
5. Friend
6. Neighbor
7. Other (Specify:_____)
9. (VOL) Refused

21b And what is their phone number?

1. Enter phone
9. Refused

22. Can I have the second person's name and address?

[USE STANDARD NAME/ADDRESS TEMPLATE]

22a. What is their relationship to you?

1. Parent
2. Brother/Sister
3. Son/Daughter
4. Other Family Member
5. Friend
6. Neighbor
7. Other (Specify:_____)
9. (VOL) Refused

22b. And what is their phone number?

1. Enter phone
9. Refused

Thanks so much. That is the end of the survey. In the next few weeks you'll be getting a letter from us with more details about the in-person interview. An interviewer will also call you to schedule the interview at a time that's convenient. If you have any questions about the survey you can reach out to the project directors at 877-699-4340, or email InequalitySurvey@abtassoc.com.

Would you like the contact information for the researchers or the Committee on the Use of Human Subjects in Research at Harvard University, who approved this study?

[IF YES: If you have any questions about the survey, you can call 877-699-4340 or email [InequalitySurvey@abtassoc.com]. Additionally, this research has been reviewed by the Committee on the Use of Human Subjects in Research at Harvard University. They can be reached at 617-496-2847, 1350 Massachusetts Avenue, 9th Floor, Suite 935, Cambridge, MA 02138 and cuhs@harvard.edu.]

QCLASS: (DUMMY PUNCH, DO NOT ASK, USE THIS FOR QUOTA)

1. Eligible, agrees to next interview [COMPLETES FULL INTERVIEW]
2. Eligible, refuses next interview [REFUSAL AT Q15 OR Q16]
3. Not Eligible [Q8>1]

Thank you very much for your assistance. Have a nice day!

D.5 In-person Interview Instrument

24256 Understanding Inequality Survey
CAPI Instrument
November 27, 2017

Hello, my name is [ ]. May I please speak with _____?

Thank you for taking the time to talk with me today. I work for Abt Associates. Abt Associates is an independent research company and we are helping Harvard EdLabs with its Understanding Inequality Survey. We are conducting interviews with people who agreed to be in this study. You might remember completing a telephone interview on [TELEPHONE INTERVIEW DATE].

This interview will include questions on your background, behaviors, childhood experiences, attitudes and personality. It will take about 2 hours to complete. When we are done, you will be paid a $150 Visa gift card, as a way of saying thank you for your time.

Your participation in this study will help us understand how to develop policies that assist in social mobility and better understand the characteristics of people living in your neighborhood.

Before we begin the survey, I would like to assure you that all of your responses on this survey will be kept private; your name will not appear in any written reports we produce. Your responses to these questions are completely voluntary. That means you may choose not to answer any question, or you may stop the interview if you wish, but we hope you don't. Some of the questions could be triggering to individuals with a history of post-traumatic stress disorder, or survivors of abuse and violence. You may find these questions upsetting. Please feel free to ask to skip these questions or discontinue the interview without penalty. You will not need to explain why you do not wish to answer or stop the interview. We will also provide you with a list of resources to local support organizations if you wish.

The information you provide will be kept private and only used for this study. By participating in this study, you will be greatly helping us further our understanding of the reasons some people do better in life and others do not.

Would you like the contact information for the researchers or the Committee on the Use of Human Subjects in Research at Harvard University, who approved this study?

[IF YES: If you have any questions about the survey, you can call 1-877-699-4340 or email InequalitySurvey@abtassoc.com. Additionally, this research has been reviewed by the Committee on the Use of Human Subjects in Research at Harvard University. They can be reached at 617-496-2847, 1350 Massachusetts Avenue, 9th Floor, Suite 935, Cambridge, MA 02138, or cuhs@harvard.edu.]

Do you have any questions before we begin?

USE STANDARD ROC/DISPOS

Before we get started, I just want to give you this copy of your rights as a participant and I also have a list of resources that may be useful.

[INTERVIEWER: PROVIDE INFORMED CONSENT AND RESOURCES LIST TO RESPONDENT]

Now, if you don't have any questions, we can get started.

Adult Characteristics

AUTO-FILL RACE, GENDER, ETC. FROM SCREENER

Base: All

Q1 What is your religious preference: is it Protestant, Catholic, Jewish, some other religion or do you have no preference?

DO NOT READ RESPONSES, CODE ALL THAT APPLY

[PROGRAMMER: ACCEPT MULTIPLE RESPONSES]

1 Protestant (Go to Q1a)
2 Catholic (Go to Q2)
3 Jewish (Go to Q2)
4 Mormon (Go to Q2)
5 Greek Orthodox (Go to Q2)
6 Russian Orthodox (Go to Q2)
7 Muslim (Go to Q2)
8 Buddhist (Go to Q2)
9 Hindu (Go to Q2)
10 Jehovah's Witness (Go to Q2)
11 Atheist (Go to Q2)
12 Agnostic (Go to Q2)
13 Something else (Go to Q2)
14 No preference (Go to Q2)
15 Nothing in particular (Go to Q2)
96 No other mentions (Go to Q2)
97 Refused (Go to Q2)
98 Don't know (Go to Q2)

Base: If Q1 = 1

Q1a What denomination do you identify with?

DO NOT READ RESPONSES, CODE ALL THAT APPLY

[PROGRAMMER: ACCEPT MULTIPLE RESPONSES]

1 Adventist (Seventh Day Adventist)
2 Anglican/Episcopalian
3 Baptist (Southern Baptist Convention, Independent Baptist, National Baptist Convention)
4 Congregationalist (United Church of Christ)
5 Holiness (Church of the Nazarene, Free Methodist Church)
6 Lutheran (Evangelical Lutheran Church in America, Lutheran Church Missouri Synod)
7 Methodist (United Methodist Church, African Methodist Episcopal)
8 Nondenominational / Ecumenical
9 Pentecostal (Assemblies of God, Church of God in Christ, Church of God)
10 Presbyterian
11 Restorationist (Church of Christ, Disciples of Christ)
12 Something else (specify)
13 No preference
14 None in particular
96 No other mentions
97 (VOL) Refused
98 (VOL) Don't know

Base: All

Q2 Aside from weddings and funerals, how often do you attend religious services?

1 More than once a week
2 Once a week
3 Once or twice a month
4 A few times a year
5 Seldom
6 Never
7 (VOL) Refused
8 (VOL) Don't know

Base: All

Q3 What is your marital status?

1 Now married
2 Widowed
3 Divorced
4 Separated
5 Never married
7 (VOL) Refused
8 (VOL) Don't know

Base: All

Q4 How many children do you have, including those no longer living with you?

NUMERIC ENTRY. INTEGERS ONLY. RANGE = -2 TO 20

-2 (VOL) Don't know
-1 (VOL) Refused

Current Household Listing

Base: All

Q5a How many people currently live in your household, including yourself? Please also include any children, under the age of 18.

NUMERIC ENTRY. INTEGERS ONLY. RANGE = -2 TO 15

-2 (VOL) Don't know
-1 (VOL) Refused

IF Q5a > 1 GO TO STATEMENT BEFORE Q5b THEN LOOP THROUGH Q5b AND Q5c WHERE THERE ARE Q5a-1 LOOPS

IF Q5a = -1, -2, OR 1 GO TO STATEMENT BEFORE Q6

Read text where Q5a > 1

I am going to ask you questions about every person who lives in your household—except you—starting with the oldest person.

Base: Q5a > 1

Q5b How is the [1ST LOOP: oldest / OTHER LOOPS: next oldest] person in your household related to you?

DO NOT READ RESPONSES

1 Mother/Stepmother (Go to Q5c)
2 Father/Stepfather (Go to Q5c)
3 Spouse (Go to Q5c)
4 Sibling (Go to Q5c)
5 Birth Child (Go to Q5c)
6 Step Child (Go to Q5c)
7 Adopted Child (Go to Q5c)
8 Foster Child (Go to Q5c)
9 Grandchild (Go to Q5c)
10 Child (not specified) (Go to Q5c)
11 Niece/nephew (Go to Q5c)
12 Aunt/uncle (Go to Q5c)
13 Cousin (Go to Q5c)
14 Grandparent (Go to Q5c)
~~15 DELETED~~
16 In-law (Go to Q5c)
17 Other relative (Go to Q5c)

18 Boyfriend/girlfriend/fiancé/fiancée (Go to Q5c)
19 Friend, not a relative (Go to Q5c)
20 Live-in aide (Go to Q5c)
21 Other unrelated person (Go to Q5c)
22 Step-sibling (Go to Q5c)
23 Other (Specify: TEXT ENTRY) (Go to Q5c)
97 Refused (Go to next loop iteration or statement before Q6)
98 Don't Know (Go to next loop iteration or statement before Q6)

Q5c How old is your [FILL FROM Q5b]?

NUMERIC ENTRY. RANGE = -2 TO 99.

-2 (VOL) Don't know
-1 (VOL) Refused

If CATI_Q11 = 6 or 9 go to Q9
If CATI_Q11 ≠ 6 or 9 go to text below

Education and Training

Show text if CATI_Q11 ≠ 6 or 9
Now I'd like to talk about your educational background.

If CATI_Q11 = 1, 2, 5, or 7 go to Q7
If CATI_Q11 = 3 or 4 go to Q6
If CATI_Q11 = 8 go to Q8b

Base: If CATI_Q11 = 3 or 4
Q6 Do you have a high school diploma or a GED?

DO NOT READ RESPONSES

1 GED (Go to Q7)
2 High School Diploma (Go to Q9 if CATI_Q11=3; Go to Q7 if CATI_Q11=4)
3 Both (Go to Q7)
4 Neither (Go to Q7)
7 Refused (Go to Q7)
8 Don't Know (Go to Q7)

Base: If CATI_Q11 = 1, 2, 4, 5, 7 or Q6 = 1, 3, 4, 7, 8

Q7 What is the highest grade or year of regular school that you have completed and gotten credit for?

DO NOT READ RESPONSES

[DISPLAY ROWS AS FOLLOWS:

IF CATI_Q11 = 1: DISPLAY 1 TO 10, 97 TO 98
IF CATI_Q11 = 2: DISPLAY 11 TO 14, 97 TO 98
IF Q6 = 1,3,4,7,8: DISPLAY 1 TO 18, 97 TO 98
IF CATI_Q11 = 4: DISPLAY 15 TO 18, 97 TO 98
IF CATI_Q11 = 5: DISPLAY 15 TO 18, 97 TO 98]

1 No formal education (Go to Q9)
2 Kindergarten (Go to Q9)
3 Grade 1 (Go to Q9)
4 Grade 2 (Go to Q9)
5 Grade 3 (Go to Q9)
6 Grade 4 (Go to Q9)
7 Grade 5 (Go to Q9)
8 Grade 6 (Go to Q9)
9 Grade 7 (Go to Q9)
10 Grade 8 (Go to Q9)
11 Grade 9 (Go to Q9)
12 Grade 10 (Go to Q9)
13 Grade 11 (Go to Q9)
14 Grade 12 (Go to Q9)
15 1st year of college (If CATI_Q11=5 go to Q8a; else go to Q9)
16 2nd year of college (If CATI_Q11=5 go to Q8a; else go to Q9)
17 3rd year of college (If CATI_Q11=5 go to Q8a; else go to Q9)
18 4th year of college (If CATI_Q11=5 go to Q8a; else go to Q9)
19 1st year of graduate/professional school (Go to Q9)
20 2nd year of graduate/professional school (Go to Q9)
21 3rd year of graduate/professional school (Go to Q9)
22 4th year of graduate/professional school (Go to Q9)
23 5th year of graduate/professional school (Go to Q9)
24 6th year of graduate/professional school (Go to Q9)
25 7th year of graduate/professional school (Go to Q9)
26 8th year of graduate/professional school (Go to Q9)
27 9th year of graduate/professional school (Go to Q9)
28 10th year of graduate/professional school (Go to Q9)
97 Refused (If CATI_Q11=5 Go to Q8a; else go to Q9)
98 Don’t know (If CATI_Q11=5 Go to Q8a; else go to Q9)

Base: If CATI_Q11 = 5

Q8a Is your Associate degree academic or is it for an occupation or vocation?

DO NOT READ RESPONSES

1 Academic (e.g., English, Math) (Go to Q9)
2 Occupation or vocation (e.g., veterinary assistant, dental hygienist) (Go to Q9)
7 Refused (Go to Q9)
8 Don't know (Go to Q9)

Base: If CATI_Q11 = 8

Q8b What is the highest degree you have received?

DO NOT READ RESPONSES

1 Master's degree (e.g., MA, MS, MBA)
2 Professional school degree (e.g., MD, JD, DDS, DVM)
3 Doctorate degree (e.g., Ph.D., Ed.D., DBA)
7 Refused
8 Don't know

Base: All

Q9 Are you currently participating in any regular schooling or in some type of training program that lasts at least two weeks, and that is designed to help you find a job, improve your job skills, or learn a new job?

DO NOT READ RESPONSES

1 Yes (Go to Q9a)
2 No (Go to Q10)
7 Refused
8 Don't know

Base: If Q9 = 1

Q9a What kind of schooling or training is that?

DO NOT READ RESPONSES

1 Regular schooling
2 General Equivalency Diploma (GED)
3 English as a Second Language
4 Computer training
5 Work study program
6 Certification or training in a health care field
7 Job search
8 Hospitality program
9 Auto repair
10 Childcare or education
11 Driving
12 Cosmetology
13 Remedial life skills

14 Accounting/Financial
15 Law/paralegal
16 Social Work
17 Construction/maintenance
18 Business/Management/Entrepreneurial
19 On the job training
20 Basic Job training
21 Certification in criminal justice
22 Other
97 Refused
98 Don't know

Base: All

Q10 Do you qualify for any educational benefits such as tuition assistance through the Post-9/11 GI Bill?

DO NOT READ RESPONSES

1 Yes
2 No
7 Refused
8 Don't know

Base: All

Q11 Have you ever served on active duty in the U.S. Armed Forces, Reserves or National Guard?

1 Never served in the military
2 Only on active duty for training in the Reserves or National Guard
3 Now on active duty
4 On active duty in the past, but not now
7 (VOL) Refused
8 (VOL) Don't know

Employment

Now I'd like to ask you a few questions about any jobs you may have.

Base: All

Q13 Are you now employed full-time, part-time or not employed?

1 EMPLOYED FULL-TIME/PART-TIME (CONTINUE TO Q14)
2 NOT EMPLOYED (SKIP TO Q22)
97 (VOL) REFUSED (SKIP TO Q23)
98 (VOL) DON'T KNOW (SKIP TO Q23)

Base: If Q13 = 1

Q14 Last week, did you have more than one job, including part-time and weekend work?

DO NOT READ RESPONSES

1 Yes
2 No
7 Refused
8 Don't know

Base: If Q13 = 1

Q15 How many hours per week do you usually work at your main job? By main job, we mean the one at which you usually work the most hours.

NUMERIC ENTRY. INTEGERS ONLY. RANGE = -3 TO 168.

-2 (VOL) Don't know
-1 (VOL) Refused
-3 (VOL) Hours vary from week to week

Base: If Q13 = 1

Q16 What kind of business or industry is this? (IF NEEDED: What do they make or do where you work?)

TEXT ENTRY

-2 (VOL) Don't know
-1 (VOL) Refused

Base: If Q13 = 1

Q17 Is this business or organization mainly manufacturing, retail trade, selling things to other businesses, or something else?

DO NOT READ RESPONSES

1 Manufacturing
2 Retail Trade
3 Wholesale Trade
4 Something Else (Specify: [TEXT ENTRY])
7 Refused
8 Don't know

Base: If Q13 = 1

Q18 What kind of work do you do, that is, what is your occupation? For example, plumber, teacher, farmer.

TEXT ENTRY

-2 (VOL) Don't know
-1 (VOL) Refused

Base: If Q13 = 1

Q19 Including overtime pay, tips, and commissions, what are your usual earnings on your primary job, before taxes or other deductions?

PROGRAMMER: NUMERIC ENTRY. ALLOW 2 DECIMAL POINTS. RANGE -2 TO 999,999.00

-2 (VOL) Don't know (Go to Q21)
-1 (VOL) Refused (Go to Q21)

Base: If Q13 = 1

Q20 Is this hourly, weekly, bi-weekly, monthly or annually?

DO NOT READ RESPONSES

1 Hourly
2 Weekly
3 Bi-weekly
4 Monthly
5 Annually
6 Other (Specify: [TEXT ENTRY])

Base: If Q13 = 1

Q21 Now I'd like to ask a few questions about benefits that may be available at your job. Through your employer are you eligible for any of the following benefits? By eligible we mean the benefit is available for you now, even if you have decided to not receive it or have not needed it.

[PROGRAMMER: Loop through the following items]

1 Health insurance
2 Sick leave
3 Paid vacation

[PROGRAMMER: Response items for each item]

DO NOT READ RESPONSES

1 Yes (Go to Q37)
2 No (Go to Q37)
7 Refused (Go to Q37)
8 Don't know (Go to Q37)

Base: If Q13 = 2

Q22 What is the main reason that you did not work for pay last week?

DO NOT READ RESPONSES

1 Retired
2 Disabled
3 Unable to Work
4 Has Job But Temporarily Absent
5 Couldn't Find Any Work
6 Child Care Problems
7 Family Responsibilities
8 In School or Other Training
9 Waiting For a New Job to Begin
10 Unemployed/laid off
11 Pregnant
12 Caring for Sick
13 No education/skills
14 Volunteer Work
15 No job
16 No work permit
17 Seasonal employment
18 Transportation problem
19 Fired
20 Moving houses
21 Don't want work
22 Quit
23 Has baby
24 Other
97 Refused
98 Don't know

Base: If Q13>1

Q23 Do you currently want a job, either full-time or part-time?

DO NOT READ RESPONSES

1 Yes/Maybe/It Depends (Go to Q23a)
2 No (Go to Q24)
3 Retired (Go to Q24)
4 Disabled (Go to Q24)
5 Unable to Work (Go to Q24)
7 Refused (Go to Q24)
8 Don't know (Go to Q24)

Base: If Q23 = 1

Q23a Have you been doing anything to find work during the past four weeks?

DO NOT READ RESPONSES

1 Yes (Go to Q23b)
2 No (Go to Q24)
7 Refused (Go to Q24)
8 Don't know (Go to Q24)

Base: If Q23a = 1

Q23b What are all the things you have done to find work during the past four weeks?

[PROGRAMMER: MULTIPLE SELECT]

DO NOT READ RESPONSES. ENTER ALL THAT APPLY.

1 Contacted Employer(s)
2 Contacted Public Employment Agency Programs/Courses
3 Contacted Private Employment Agency
4 Contacted Friends or Relatives
5 Interviewed for a Job
6 Contacted School/University Employer Center
7 Sent Out Resumes/Filled out Applications
8 Checked Union/Professional Registers
9 Placed or Answered Ads
10 Looked at Ads Directly
11 Attended Job Training
12 Nothing
13 Other
96 No other mentions
97 Refused
98 Don't know

Base: If Q13 > 1

Q24 Last week, could you have started a job if one had been offered?

DO NOT READ RESPONSES

1 Yes (Includes "maybe" under conditions on hours, wages or type of work) (Go to Q37)
2 No (Go to Q24a)
7 Refused (Go to Q37)
8 Don't know (Go to Q37)

Base: If Q24=2

Q24a Why is that?

[PROGRAMMER: MULTIPLE SELECT]

DO NOT READ RESPONSES. ENTER ALL THAT APPLY

1 Waiting for a new job to begin
2 Own temporary illness
3 Going to school
4 Childcare responsibilities
5 Family responsibilities
6 Transport problems
7 Moving houses
8 Mental/Physical Illness
9 Other
96 No other mentions
97 Refused
98 Don't know

Base: All

Q37 What kind of health insurance or health care coverage do you have for yourself?

[INTERVIEWER: RESPONSE SHOULD BE THINGS LIKE "THROUGH MY EMPLOYER" OR "MEDICAID" WE DON'T NEED NAME OF INSURANCE COMPANY]

TEXT ENTRY

-2 (VOL) Don't know
-1 (VOL) Refused

Adult Income and Benefits

Base: All

BEN1 Next I'd like to talk with you about any income or public assistance you ((IF (Q5b = 5 or Q5b = 6 or Q5b = 7 or Q5b = 8 or Q5b = 10) and Q5c < 18 for any HH member rostered: or your child under the age of 18) may receive. Do you receive any form of public assistance or benefits from the government?

DO NOT READ RESPONSES

1 Yes (Go to Q36)
2 No (Go to INC1)
7 Refused (Go to Q36)
8 Don't know (Go to Q36)

Q36 Are you (IF (Q5b = 5 or Q5b = 6 or Q5b = 7 or Q5b = 8 or Q5b = 10) and Q5c < 18 for any HH member rostered: or your child under the age of 18) now receiving help from the Supplemental Security Income program, called SSI?

DO NOT READ RESPONSES

1 Yes (IF Q5b = 5 or Q5b = 6 or Q5b = 7 or Q5b = 8 or Q5b = 10 go to Q36a;
2 No (Go to Q38)
7 Refused (Go to Q38)
8 Don't know (Go to Q38)

Base: If (Q5b = 5 or Q5b = 6 or Q5b = 7 or Q5b = 8 or Q5b = 10) and Q36 = 1
Q36a Is the SSI for you or for your child?

If Q5b ≠ 5 and Q5b ≠ 6 and Q5bb ≠ 7 and Q5b ≠ 8 and Q5b ≠ 10 autopunch Q36a=1
If not [{Q5b = 5 or Q5b = 6 or Q5b = 7 or Q5b = 8 or Q5b = 10} and Q5c < 18] autopunch Q36a=1

[INTERVIEWER- IF NECESSARY: READ ANSWER CHOICES ALOUD]

1 You
2 Your child
3 Both you and your child
7 REFUSED
8 DON'T KNOW

Base: IF BEN1 = 1,7,8
Q38 Now I'd like to ask you about cash assistance some families receive on a regular basis. For example, they may get a monthly check. Some people call this assistance "welfare," AFDC, TANF ("tan-eff") or "public aid." I'll use the word "welfare." Are you regularly receiving welfare benefits now?

DO NOT READ RESPONSES

1 Yes
2 No
7 Refused
8 Don't know

Base: IF BEN1 = 1,7,8

Q39 Do you (IF Q5b = 5 or Q5b = 6 or Q5b = 7 or Q5b = 8 or Q5b = 10: or your child) receive food stamps or WIC ("wick")?

DO NOT READ RESPONSES

1 Yes
2 No
7 Refused
8 Don't know

Base: If BEN1=2

INC1 What was your total annual household income last year, in 2016, before taxes?

NUMERIC ENTRY. ALLOW UP TO 2 DECIMAL POINTS. RANGE = 0 TO 999999.00.

-2 (VOL) Don't know
-1 (VOL) Refused

Base: If BEN1=1,7,8

INC_Alt1 What was your total annual household income last year, in 2016, before taxes, as you reported to the government?

NUMERIC ENTRY. ALLOW UP TO 2 DECIMAL POINTS. RANGE = 0 TO 999999.00.

-2 (VOL) Don't know
-1 (VOL) Refused

Base: IF ((INC1 = -1 or -2) or (INC_Alt1= -1 or -2))

INC1a Could you please tell me if it was… (READ LIST)

1. Under $20,000 (ASK INC1b)
2. $20,000 to under $50,000, or (ASK INC1c)
3. $50,000 or more (ASK INC1d)

-1 (VOL) Don't know
-2 (VOL) Refused

Base: If INC1a=1

INC1b Is it…(READ LIST)

1. Under $10,000, or
2. $10,000 to under $20,000?

-1 (VOL) Don't know
-2 (VOL) Refused

Base: If INC1a=2

INC1c Is it…(READ LIST)

1. $20,000 to under $30,000
2. $30,000 to under $40,000, or
3. $40,000 to under $50,000?

-1 (VOL) Don't know
-2 (VOL) Refused

Base: If INC1a=3

INC1d Is it…(READ LIST)

1. $50,000 to under $75,000
2. $75,000 to under $100,000
3. $100,000 to under $150,000, or
4. $150,000 or more?

-1 (VOL) Don't know
-2 (VOL) Refused

Base: If BEN1=2

INC2 If you are not the only income earner in your household, what was your individual annual income last year, in 2016, before taxes?

NUMERIC ENTRY. ALLOW UP TO 2 DECIMAL POINTS. RANGE = -3 TO 999999.00.

-3 (VOL) Only income earner in household
-2 (VOL) Don't know
-1 (VOL) Refused

Base: If BEN1=1,7,8

INC_Alt2 If you are not the only income earner in your household, what was your individual annual income last year, in 2016, before taxes, as you reported to the government?

NUMERIC ENTRY. ALLOW UP TO 2 DECIMAL POINTS. RANGE = -3 TO 999999.00.

-3 (VOL) Only income earner in household
-2 (VOL) Don't know
-1 (VOL) Refused

Base: IF ((INC2 = -1 or -2) or (INC_Alt2= -1 or -2))

INC2a Could you please tell me if it was… (READ LIST)

1. Under $20,000 (ASK INC1b)
2. $20,000 to under $50,000, or (ASK INC1c)
3. $50,000 or more (ASK INC1d)

-1 (VOL) Don't know

-2 (VOL) Refused

Base: If INC2a=1

INC2b Is it…(READ LIST)

1. Under $10,000, or
2. $10,000 to under $20,000?

-1 (VOL) Don't know

-2 (VOL) Refused

Base: If INC2a=2

INC2c Is it…(READ LIST)

1. $20,000 to under $30,000
2. $30,000 to under $40,000, or
3. $40,000 to under $50,000?

-1 (VOL) Don't know

-2 (VOL) Refused

Base: If INC2a=3

INC2d Is it…(READ LIST)

1. $50,000 to under $75,000
2. $75,000 to under $100,000
3. $100,000 to under $150,000, or
4. $150,000 or more?

-1 (VOL) Don't know

-2 (VOL) Refused

Base: All

Q25 Now I want to ask about accounts at a bank, savings and loan, or credit union. Please think about any accounts which you own, including joint accounts. Do you currently have any checking accounts, savings accounts, or any other type of bank account at a bank, credit union, or other financial institution? (IF NEEDED: Do not include retirement savings like 401(k), 403(b), and IRAs.)

DO NOT READ RESPONSES

1 Yes (Go to Q25a)

2 No (Go to Q25b)

7 Refused (Go to INC3)

8 Don't know (Go to INC3)

Base: If Q25 = 1

Q25a About how much is in these accounts all together?

PROGRAMMER: NUMERIC ENTRY. ALLOW UP TO 2 DECIMAL POINTS. RANGE = -2 TO 999999.00. (Go to INC1)

-2 (VOL) Don't know (Go to INC3)
-1 (VOL) Refused (Go to INC3)

Base: If Q25 = 2

Q25b What is the most important reason you don't have a bank account?

DO NOT READ RESPONSES

1 Don't write enough checks to make it worthwhile
2 Do not have enough money to keep in an account
3 Don't like dealing with banks
4 Service charges are too high
5 No bank has convenient hours or locations
6 Institutions will take their money due to debt
7 Bank closed account due to overdraft fees
8 Don't have enough money
9 Other response
97 Refused
98 Don't know

Base: All

INC3 In your own words, what would you say are the main factors that have contributed to your current financial situation? (PROBE: Up to 3 mentions; accept multiple responses)

Response 1 [TEXT ENTRY]
Response 2 [TEXT ENTRY]
Response 3 [TEXT ENTRY]

INC3DK (VOL) Don't know
INC3RF (VOL) Refused

Base: All

Q26 Do you or does anyone in your household own a car or truck, or other motor vehicle that runs and can be driven on the road?

DO NOT READ RESPONSES

1 Yes (Go to Q26a)
2 No (Go to Q27)
7 Refused (Go to Q27)
8 Don't know (Go to Q27)

Base: If Q26 = 1

Q26a Thinking about the vehicles that you own, did you borrow money or get financing to purchase any of your vehicles?

DO NOT READ RESPONSES

1 Yes
2 No
7 Refused
8 Don't know

Base: All

Q27 Do you or does anyone in your household have any unpaid medical bills?

DO NOT READ RESPONSES

1 Yes (Go to Q27a)
2 No (Go to Q28)
7 Refused (Go to Q28)
8 Don't know (Go to Q28)

Base: If Q27 = 1

Q27a About how much do you and your household still owe on your medical bills?

NUMERIC ENTRY. ALLOW UP TO 2 DECIMAL POINTS. RANGE = -2 TO 999999.00

-2 (VOL) Don't know
-1 (VOL) Refused

Base: All

Q28 Do you have any credit or charge cards, including major credit cards like Visa or MasterCard, or charge cards from a store or gas station such as Sears or Mobil?

DO NOT READ RESPONSES

1 Yes (Go to Q28a)
2 No (Go to Q29)
7 Refused (Go to Q29)
8 Don't know (Go to Q29)

Base: If Q28 = 1

Q28a About how much do you and your household currently owe on all your credit and charge cards?

PROGRAMMER: NUMERIC ENTRY. ALLOW UP TO 2 DECIMAL POINTS. RANGE -2 TO 999999.00.

-2 (VOL) Don't know
-1 (VOL) Refused

Base: All

Q29 Are debt collectors contacting you to ask that you make payments on any debts? (If needed: Not only your credit cards.)

DO NOT READ RESPONSES

1 Yes
2 No
7 Refused
8 Don't know

Base: All

Q30 Do you have an overdue balance you're paying down on your utilities?

DO NOT READ RESPONSES

1 Yes
2 No
7 Refused
8 Don't know

Base: All

Q31 Are there any bills that you are ignoring or not making any payment on?

DO NOT READ RESPONSES

1 Yes
2 No
7 Refused
8 Don't know

Base: All

Q32 If you needed to borrow $500 for three months, is there some person or place you could borrow it from?

DO NOT READ RESPONSES

1 Yes (Go to Q32a)
2 No (Go to Q33)
3 Would not borrow (Go to Q33)
7 Refused (Go to Q33)
8 Don't know (Go to Q33)

Base: If Q32 = 1

Q32a Where would you go first to borrow $500 for three months?

DO NOT READ RESPONSES

1 Friends or family
2 A finance company
3 A payday loan at a check cashing outlet
4 Someone in my neighborhood who lends out money and charges interest
5 A community loan fund (or church loan fund)
6 A cash advance on my credit card
7 A bank (or savings bank, savings & loan, or credit union)
8 A pawn shop
9 A furniture store
10 Job/pension/union
11 I would not borrow
12 Other
97 Refused
98 Don't know

Base: All

Q33 How often do you or your household put off buying something you need because you don't have money? Would you say all the time, frequently, occasionally, rarely, or never?

DO NOT READ RESPONSES

1 All the time
2 Frequently
3 Occasionally
4 Rarely
5 Never
7 Refused
8 Don't know

If CATI_Q1A > 1 go to Q34
If CATI_Q1A = 1 go to Q35

Base: CATI_Q1A > 1

Q34 Did you ever live with your mother for a period of 6 months or more, at age 25 or older?

NOTE TO INTERVIEWER: MOTHER MAY BE BIOLOGICAL OR STEPMOTHER OR ADOPTIVE MOTHER

DO NOT READ RESPONSES

1 Yes (Go to Q34a)
2 No (Go to Q35)
6 No mother (Go to LT1)
97 Refused (Go to Q35)
98 Don't know (Go to Q35)

Base: If Q34 = 1

Q34a Would you say that living with her was mainly to help your mother out, to help you out, or because it would be helpful to both of you?

DO NOT READ RESPONSES

1 To help your mother out
2 To help you out
3 Helpful to both
4 Neither
7 Refused
8 Don't know

Base: All

Q35 Did any of your brothers or sisters ever live with your mother for a period of 6 months or more, at age 25 or older?

NOTE TO INTERVIEWER: SIBLNGS MAY BE BIOLOGICAL OR STEP-SIBLINGS OR ADOPTED SIBLINGS OR HALF SIBLINGS

DO NOT READ RESPONSES

1 Yes (Go to Q35a)
2 No (Go to LT1)
6 No brothers or sisters / Brothers or sisters all younger than age 25 (Go to LT1)
7 Refused (Go to LT1)
8 Don't know (Go to LT1)

Base: If Q35 = 1

Q35a Would you say that your brother or sister living with your mother was mainly to help your mother out, to help your brother or sister out, or because it would be helpful to both of them?

DO NOT READ RESPONSES

1 To help your mother out
2 To help your brother or sister out
3 Helpful to both
4 Neither
97 Refused
98 Don't know

Lifetime Traumas

Base: All

LT1 Now I'd like to ask you a few questions about events you may have experienced. For each of the following events, please indicate whether the event occurred at any point in your life…

[PROGRAMMER: Loop through following items]

1 Have you ever been homeless?
2 Have you ever been in a major fire, flood, earthquake, or other natural disaster?
3 Did you ever have a life-threatening illness or accident?
4 Did your spouse or a child of yours ever have a life-threatening illness or accident?
5 Have you ever fired a weapon in combat or been fired upon in combat?
6 Has your spouse, partner, or child ever been addicted to drugs or alcohol?
7 Were you the victim of a serious physical attack or assault in your life?
8 Has a child of yours ever died?

[PROGRAMMER: Response options for each item.]

DO NOT READ RESPONSES

1 Yes
2 No
7 Refused
8 Don't know

Base: All

LT2 Have you ever been arrested by the police or taken into custody for an illegal or delinquent offense? Do not include arrests for minor traffic violations.

DO NOT READ RESPONSES

1 Yes (Go to LT3)
2 No (Go to LT4)
7 Refused (Go to LT4)
8 Don't know (Go to LT4)

Base: If LT2 = 1

LT3 In total, how many times have you been arrested?

NUMERIC ENTRY. INTEGERS ONLY. RANGE = -2, -1, 1 TO 999.

-2 (VOL) Don't know
-1 (VOL) Refused

Lifetime Traumas Before the Age of 18

Base: All

LT4 Before you were 18 years old, did you ever have to do a year of school over again?

DO NOT READ RESPONSES

1 Yes
2 No
7 Refused
8 Don't know

Base: All

LT5 Before you were 18 years old, were you ever in trouble with the police?

DO NOT READ RESPONSES

1 Yes
2 No
7 Refused
8 Don't know

Base: All

LT6 Before you were 18 years old, did either of your parents drink or use drugs so often that it caused problems in the family?

DO NOT READ RESPONSES

1 Yes
2 No
7 Refused
8 Don't know

Base: All

LT7 Before you were 18 years old, were you ever physically abused by either of your parents?

DO NOT READ RESPONSES

1 Yes
2 No
7 Refused
8 Don't know

Adult Health

Show text to all
Now I'd like to ask you some questions about your general health.

Base: All
HEAL1 In general, how is your health: excellent, very good, good, fair, or poor?

DO NOT READ RESPONSES

1 Excellent
2 Very good
3 Good
4 Fair
5 Poor
7 Refused
8 Don't know

Base: All
HEAL2 In general, how has your health been in the past month: excellent, very good, good, fair, or poor?

DO NOT READ RESPONSES

1 Excellent
2 Very good
3 Good
4 Fair
5 Poor
7 Refused
8 Don't know

Base: All
HEAL3 In the past year, have you had a routine physical examination?

DO NOT READ RESPONSES

1 Yes
2 No
7 Refused
8 Don't know

Base: All
HEAL4 How tall are you?

If respondent only knows height in meters:

5 feet 0 inches = 1.52 meters / 152 cm
5 feet 1 inches = 1.55 meters / 155 cm
5 feet 2 inches = 1.57 meters / 157 cm
5 feet 3 inches = 1.60 meters / 160 cm
5 feet 4 inches = 1.63 meters / 163 cm
5 feet 5 inches = 1.65 meters / 165 cm
5 feet 6 inches = 1.68 meters / 168 cm
5 feet 7 inches = 1.70 meters / 170 cm
5 feet 8 inches = 1.73 meters / 173 cm
5 feet 9 inches = 1.75 meters / 175 cm
5 feet 10 inches = 1.78 meters / 178 cm
5 feet 11 inches = 1.80 meters / 180 cm
6 feet 0 inches = 1.83 meters / 183 cm
6 feet 1 inches = 1.85 meters / 185 cm
6 feet 2 inches = 1.88 meters / 188 cm
6 feet 3 inches = 1.91 meters / 191 cm
6 feet 4 inches = 1.93 meters / 193 cm
6 feet 5 inches = 1.96 meters / 196 cm
6 feet 6 inches = 1.98 meters / 198 cm
6 feet 7 inches = 2.01 meters / 201 cm
6 feet 8 inches = 2.03 meters / 203 cm

Feet: NUMERIC ENTRY. INTEGERS ONLY. RANGE = 3 TO 8.
Inches: NUMERIC ENTRY. INTEGERS ONLY. RANGE = 0 TO 11.

HEAL4DK (VOL) Don’t know
HEAL4RF (VOL) Refused

Base: All

HEAL5 How much do you weigh? (IF NEEDED: Your best guess is fine.)

If respondent only knows weight in kilograms:

120 pounds = 54 kilograms
130 pounds = 59 kilograms
140 pounds = 64 kilograms
150 pounds = 68 kilograms
160 pounds = 73 kilograms
170 pounds = 77 kilograms
180 pounds = 82 kilograms
190 pounds = 86 kilograms
200 pounds = 91 kilograms
210 pounds = 95 kilograms
220 pounds = 100 kilograms
230 pounds = 104 kilograms
240 pounds = 109 kilograms
250 pounds = 113 kilograms

Pounds: NUMERIC ENTRY. INTEGERS ONLY. RANGE = -2 TO -1, 50 TO 600.

-2 (VOL) Don't know
-1 (VOL) Refused

Base: All

HEAL6 Have you ever suffered from depression, anxiety, emotional distress, or mental illness of any kind?

DO NOT READ RESPONSES

1 Yes (Go to HEAL7)
2 No (Go to DIS1)
7 Refused (Go to DIS1)
8 Don't know (Go to DIS1)

Base: If HEAL6 = 1
HEAL7 What was the illness?

DO NOT READ RESPONSES. ALLOW MULTIPLE RESPONSES.

[PROGRAMMER: MULTIPLE SELECT]

1 Depression
2 Anxiety
3 Bipolar
4 Schizophrenia
5 Paranoia
9 ADHD
10 PTSD
11 Eating Disorder
12 Emotional distress
13 Other
96 No other mentions
97 Refused
98 Don't know

Discrimination[1]

Now I have a few questions about discrimination. Sometimes people feel like they are discriminated against, or treated badly or differently because of their race or ethnicity.

Base: All
DIS1 Can you think of one or more occasions in the last 6 months when you felt you were treated unfairly because of your race or ethnicity in the following places?

[PROGRAMMER: Loop through following items]

1 At your school or work?
2 In a store where you were shopping or a restaurant where you wanted to eat?
3 When you met someone for the first time?
4 In dealing with the police?

[PROGRAMMER: Response options for each item.]

DO NOT READ RESPONSES

1 Yes
2 No
7 Refused
8 Don't know

Show text for all
Sometimes people feel they are discriminated against, or treated badly or differently because they might not have quite as much money as other people, or because of the way they dress or talk.

[1] From MTO.

Base: All

DIS2 Can you think of one or more occasions in the last 6 months when you felt you were treated unfairly because of how much money your family has or the way you dress or talk?

[PROGRAMMER: Loop through following items]

1 At your school or work?
2 In a store where you were shopping or a restaurant where you wanted to eat?
3 When you met someone for the first time?
4 In dealing with the police?

[PROGRAMMER: Response options for each item.]

DO NOT READ RESPONSES

1 Yes
2 No
7 Refused
8 Don't know

Discount Rates

Show text to all

Suppose that after having helped a friend with some small jobs, they offer to send you a small amount of money in return for your help. They tell you that they can either send you something now, or send you a little more if you are willing to wait one month. If they pay you now, they will put $40 in the mail tomorrow. If they pay you one month from now, they will send you slightly more than that. Suppose that you trust them to pay you what they promise, when they promise it, and that either payment is equally convenient for them.

Base: All

DSC1 Would you rather they mailed you $40 tomorrow or $47 one month from now?

DO NOT READ RESPONSES

1 $40 tomorrow (Go to DSC4)
2 $47 one month from now (Go to DSC2)
7 Refused (Go to DSC2)
8 Don't know (Go to DSC2)

Base: If DSC1 ≠ 1

DSC2 Now suppose the choice were between $40 now and $45 one month from now. Would you rather they mailed you $40 tomorrow or $45 one month from now?

DO NOT READ RESPONSES

1 $40 tomorrow (Go to statement before RISK1)
2 $45 one month from now (Go to DSC3)
7 Refused (Go to DSC3)
8 Don't know (Go to DSC3)

Base: If DSC2 ≠ 1

DSC3 Now suppose the choice were between $40 now and $42 one month from now. Would you rather they mailed you $40 tomorrow or $42 one month from now?

DO NOT READ RESPONSES

1 $40 tomorrow (Go to statement before RISK1)
2 $42 one month from now (Go to statement before RISK1)
7 Refused (Go to statement before RISK1)
8 Don't know (Go to statement before RISK1)

Base: If DSC1 = 1

DSC4 Now suppose the choice were between $40 now and $50 one month from now. Would you rather they mailed you $40 tomorrow or $50 one month from now?

DO NOT READ RESPONSES

1 $40 tomorrow (Go to DSC5)
2 $50 one month from now (Go to statement before RISK1)
7 Refused (Go to statement before RISK1)
8 Don't know (Go to statement before RISK1)

Base: If DSC4 = 1

DSC5 Now suppose the choice were between $40 now and $55 one month from now. Would you rather they mailed you $40 tomorrow or $55 one month from now?

DO NOT READ RESPONSES

1 $40 tomorrow
2 $55 one month from now
7 Refused
8 Don't know

Risk Aversion

Show text to all

Suppose you have a choice between two, equally good jobs that lasted for one month. The first would pay you $600 for the month. The second job would pay you an amount that depends on how the company as a whole did during that month. It is possibly better paying, but your earnings will be less certain.

Base: All

RISK1 There is a 50-50 chance that the second job will pay $1200, and a 50-50 chance it will pay $400. Which would you choose -- the job that pays $600 for sure, or the job with an equal chance of paying either $1200 or $400?

DO NOT READ RESPONSES

1 $600 for sure (Go to RISK4)
2 Equal chance of paying either $1200 or $400 (Go to RISK2)
7 Refused (Go to RISK2)
8 Don't know (Go to RISK2)

Base: If RISK1 ≠ 1

RISK2 Now suppose there is a 50-50 chance that the second job will pay $1200, and a 50-50 chance that it will pay $300. Which would you choose -- the job that pays $600 for sure, or the job with an equal chance of paying either $1200 or $300?

DO NOT READ RESPONSES

1 $600 for sure (Go to RES)
2 Equal chance of paying either $1200 or $300 (Go to RISK3)
7 Refused (Go to RISK3)
8 Don't know (Go to RISK3)

Base: If RISK2 ≠ 1

RISK3 Now suppose there is a 50-50 chance that the second job will pay $1200, and a 50-50 chance that it will pay $150. Which would you choose -- the job that pays $600 for sure, or a job with an equal chance of paying either $1200 or $150?

DO NOT READ RESPONSES

1 $600 for sure (Go to RES)
2 Equal chance of paying either $1200 or $150 (Go to RES)
7 Refused (Go to RES)
8 Don't know (Go to RES)

Base: If RISK1 = 1

RISK4 Now suppose there is a 50-50 chance that the second job will pay $1200, and a 50-50 chance that it will pay $480. Which would you choose -- the job that pays $600 for sure, or a job with an equal chance of paying either $1200 or $480?

DO NOT READ RESPONSES

1 $600 for sure (Go to RISK5)
2 Equal chance of paying either $1200 or $480 (Go to RES)
7 Refused (Go to RES)
8 Don't know (Go to RES)

Base: If RISK4 = 1

RISK5 Now suppose there is a 50-50 chance that the second job will pay $1200, and a 50-50 chance that it will pay $540. Which would you choose -- the job that pays $600 for sure, or a job with an equal chance of paying either $1200 or $540?

DO NOT READ RESPONSES

1 $600 for sure
2 Equal chance of paying either $1200 or $540
7 Refused
8 Don't know

Brief Resilience Scale[2]

Base: All

RES I am going to read you a series of sentences. For each of these, please tell me how much you agree with each sentence. There are no right or wrong answers.

[PROGRAMMER: Loop through the following items.]

1 I tend to bounce back quickly after hard times
2 I have a hard time making it through stressful event
3 It does not take me long to recover from a stressful event
4 It is hard for me to snap back when something bad happens
5 I usually come through difficult times with little trouble
6 I tend to take a long time to get over set-backs in my life

[PROGRAMMER: Response options for each item.]

USE SHOWCARD #1

1 Strongly disagree
2 Disagree
3 Neutral
4 Agree
5 Strongly agree
7 (VOL) Refused
8 (VOL) Don't know

[2] Smith et. al 2008

Childhood Experience

Base: All
CHHH1 Who did you live with for the majority of the time at age 9?

DO NOT READ RESPONSES, CODE ALL THAT APPLY.
INTERVIEWER IF NEEDED: FOSTER PARENT SHOULD BE NOTED IN OTHER SPECIFY.

[PROGRAMMER: ACCEPT MULTIPLE RESPONSES]

1 Mother/Stepmother
2 Father/Stepfather
3 Grandparent(s)
4 Younger brother(s)/stepbrother(s)
5 Younger sister(s)/stepsister(s)
6 Older brother(s)/stepbrother(s)
7 Older sister(s)/stepsister(s)
8 Twin/triplet/etc. brother(s)
9 Twin/triplet/etc. sister(s)
10 Niece/nephew(s)
11 Aunt/uncle(s)
12 Cousin(s)
13 Other (Specify: [TEXT BOX])
96 No other mentions
97 Refused
98 Don't know

Base: If CHHH1_4 selected.
CHHH2 How many younger brothers or stepbrothers did you live with at age 9? (If needed: For a majority of the time.)

[NUMERIC ENTRY. RANGE = -2-9]

-2 (VOL) Don't know
-1 (VOL) Refused

Base: If CHHH1_5 selected.
CHHH3 How many younger sisters or stepsisters did you live with at age 9? (If needed: For a majority of the time.)

[NUMERIC ENTRY. RANGE = -2-9]

-2 (VOL) Don't know
-1 (VOL) Refused

Base: If CHHH1_6 selected.

CHHH4 How many older brothers or stepbrothers did you live with at age 9? (If needed: For a majority of the time.)

[NUMERIC ENTRY. RANGE = -2-9]

-2 (VOL) Don't know
-1 (VOL) Refused

Base: If CHHH1_7 selected.

CHHH5 How many older sisters or stepsisters did you live with at age 9? (If needed: For a majority of the time.)

[NUMERIC ENTRY. RANGE = -2-9]

-2 (VOL) Don't know
-1 (VOL) Refused

Base: If CHHH1_8 selected.

CHHH6 How many brothers born at the same time as you did you live with at age 9? (If needed: For a majority of the time.)

[NUMERIC ENTRY. RANGE = -2-9]

-2 (VOL) Don't know
-1 (VOL) Refused

Base: If CHHH1_9 selected.

CHHH7 How many sisters born at the same time as you did you live with at age 9? (If needed: For a majority of the time.)

[NUMERIC ENTRY. RANGE = -2-9]

-2 (VOL) Don't know
-1 (VOL) Refused

Base: If CHHH1_10 selected.

CHHH8 How many nieces and nephews did you live with at age 9? (If needed: For a majority of the time.)

[NUMERIC ENTRY. RANGE = -2-9]

-2 (VOL) Don't know
-1 (VOL) Refused

Base: If CHHH1_11 selected.
CHHH9 How many uncles and aunts did you live with at age 9? (If needed: For a majority of the time.)

[NUMERIC ENTRY. RANGE = -2-9]

-2 (VOL) Don't know
-1 (VOL) Refused

Base: If CHHH1_12 selected.
CHHH10 How many cousins did you live with at age 9? (If needed: For a majority of the time.)

[NUMERIC ENTRY. RANGE = -2-9]

-2 (VOL) Don't know
-1 (VOL) Refused

Base: If CHHH1_13 selected.
CHHH11 How many [FILL WITH TEXT FROM CHHH1_13txt] did you live with at age 9? (If needed: For a majority of the time.)

[NUMERIC ENTRY. RANGE = -2-9]

-2 (VOL) Don't know
-1 (VOL) Refused

Base: All
NB1 How many times did you move between the ages of 5 and 17?

NUMERIC ENTRY. INTEGERS ONLY. RANGE = -2-12.
INTERVIEWER: ENTER 12 FOR "12 OR MORE"

0 (Go to NB4)
1-12 (Go to NB2)
-2 (VOL) Don't know (Go to NB4)
-1 (VOL) Refused (Go to NB4)

[PROGRAMMER: Construct read-in variable NB1txt from NB1 where 1 = "1st", 2 = "2nd", 3 = "3rd", 4 = "4th", 5 = "5th. ONLY PROGRAM 5 LOOPS.]

[PROGRAMMER: If NB1 = 1 to 5, loop through mentions 1 to 5 in NB2 and NB3. for each loop NB2 then NB3 should be asked before moving to the next loop iteration.]

Base: Loop iteration from NB1
NB2 How old were you when you moved the [FILL WITH NB1txt] time?

NUMERIC ENTRY, ACCEPT MULTIPLE ANSWERS. INTEGERS ONLY. RANGE = -2, -1, 5-17.

-2 (VOL) Don't know
-1 (VOL) Refused

Base: Loop iteration from NB1

NB3 What was the main reason you moved at age [INSERT AGE FROM NB2]?

TEXT ENTRY

98 (VOL) Don't know
99 (VOL) Refused

Base: All

NB4 Please think about the neighborhood that you spent the most amount of time in when you were under the age of 18. I'd like to ask some questions about that neighborhood. For the following, the answer choices are "Very safe, Safe, Unsafe, Very unsafe."

[PROGRAMMER: Loop through following items]

1 How safe were the parking lots and sidewalks near your neighborhood school?
2 How safe did you feel at home alone at night?
3 How safe were the streets near your home during the day?
4 How safe were the streets near your home at night?

[PROGRAMMER: Response options for each item]

USE SHOWCARD #2

1 Very safe
2 Safe
3 Unsafe
4 Very unsafe
7 (VOL) Refused
8 (VOL) Don't know

Base: All

NB5 The next questions ask about problems in your neighborhood. Your answer choices are: Big problem, Small problem, No problem at all. In your neighborhood, how bad of a problem was...

[PROGRAMMER: Loop through following items]

1 Litter or trash on the streets or sidewalks?
2 Graffiti or writing on the walls?
3 People drinking in public?

[PROGRAMMER: Response options for each item]

USE SHOWCARD #3

1 Big problem
2 Small problem
3 No problem at all
7 (VOL) Refused
8 (VOL) Don't know

Base: All

NB6 Can you remember any of the following happening in your neighborhood?

[PROGRAMMER: Loop through following items]

1 Was anyone's purse, wallet, or jewelry snatched from them?
2 Was anyone threatened with a knife or gun?
3 Was anyone beaten or assaulted?
4 Was anyone stabbed or shot?
5 Did anyone try to break into your home?

[PROGRAMMER: Response options for each item]

1 Yes
2 No
7 (VOL) Refused
8 (VOL) Don't know

Base: All

NB7 If a group of neighborhood children were skipping school and hanging out on a street corner, how likely is it that your neighbors would do something about it? Would you say very likely, likely, unsure, unlikely, or very unlikely?

DO NOT READ RESPONSES

1 Very likely
2 Likely
3 Unsure (includes don't know)
4 Unlikely
5 Very unlikely
7 Refused

Base: All

NB8 If some children were spray-painting graffiti on a local building, how likely is it that your neighbors would do something about it? Would you say very likely, likely, unsure, unlikely, or very unlikely?

DO NOT READ RESPONSES

1 Very likely
2 Likely
3 Unsure (includes don't know)
4 Unlikely
5 Very unlikely
7 Refused

Finances[3]

Read to all
Now think about your family when you were growing up, from birth until age 16.

Base: All
CFIN1 Would you say your family during that time was pretty well off financially, about average, or poor?

DO NOT READ RESPONSES

1 Pretty well off
2 About average
3 Poor
7 Refused
8 Don't know

Base: All
CFIN2 While you were growing up, from birth until age 16, did financial difficulties ever cause you or your family to move to a different place?

DO NOT READ RESPONSES

1 Yes (Go to CFIN2a)
2 No (Go to CFIN3)
7 Refused (Go to CFIN3)
8 Don't know (Go to CFIN3)

Base: If CFIN2 = 1
CFIN2a How old were you at the time?

NUMERIC ENTRY. INTEGERS ONLY. RANGE = -2 TO 16.

-2 (VOL) Don't know
-1 (VOL) Refused

Base: All
CFIN3 From birth until age 16, was there a time when you or your family received help from relatives because of financial difficulties?

DO NOT READ RESPONSES

1 Yes (Go to CFIN3a)
2 No (Go to CFIN4)
7 Refused (Go to CFIN4)
8 Don't know (Go to CFIN4)

[3] Mostly HRS, some MTO Baseline,

Base: If CFIN3 = 1
CFIN3a How old were you at the time?

NUMERIC ENTRY. INTEGERS ONLY. RANGE = -2 TO 16.

-2 (VOL) Don't know
-1 (VOL) Refused

Base: All
CFIN4 From birth until age 16, was there a time of several months or more when your father had no job?

DO NOT READ RESPONSES

1 Yes (Go to CFIN4a)
2 No (Go to CFIN5)
6 Did not have father / Father deceased (Go to CFIN6)
8 Refused (Go to CFIN5)
7 Don't know (Go to CFIN5)

Base: If CFIN4 = 1
CFIN4a How old were you at the time?

NUMERIC ENTRY. INTEGERS ONLY. RANGE = -2 TO 16.

-2 (VOL) Don't know
-1 (VOL) Refused

Base: CFIN4≠6
CFIN5 What was your father's occupation when you were age 16?

TEXT ENTRY

-4 (VOL) Father was unemployed
-3 (VOL) Did not have father / Father deceased
-2 (VOL) Don't know
-1 (VOL) Refused

Base: All
CFIN6 Did your mother ever get AFDC (IF NEEDED: "Aid to Families with Dependent Children") or welfare from your birth until age 16?

DO NOT READ RESPONSES

1 Yes (Go to CFIN6a)
2 No (Go to CFIN7)
6 Did not have mother / Mother deceased (Go to CFIN8)
7 Refused (Go to CFIN7)
8 Don't know (Go to CFIN7)

Base: If CFIN6 = 1
CFIN6a How old were you at the time?

NUMERIC ENTRY. INTEGERS ONLY. RANGE = -2 TO 16.

7 (VOL) Refused
8 (VOL) Don't know

PROGRAMMER: SKIP CFIN7 IF CFIN4=6 OR CFIN6=6
Base: IF CFIN4≠6 AND CFIN6≠6
CFIN7 Did you live with both of your parents until you were 16?

DO NOT READ RESPONSES

1 Yes
2 No (includes one or both parents deceased / some of the time)
7 Refused
8 Don't know

Base: All
CFIN8 Did you ever live in the same household with a grandparent for a year or more until you were 16?

DO NOT READ RESPONSES

1 Yes
2 No
7 Refused
8 Don't know

PROGRAMMER: SKIP CFIN9 IF CFIN6=6
Base: CFIN6≠6
CFIN9 What portion of the time did your mother work outside the home until you were 16: all of the time, some of the time, or not at all?

DO NOT READ RESPONSES

1 All of the time
2 Some of the time
3 Not at all
4 Mother deceased or not around
7 Refused
8 Don't know

Base: All

CFIN10 How often do you remember the following happening to your parents or guardian until you were age 16? Would you say always, often, sometimes, rarely, or never? [INSERT STATEMENT FROM LOOP SHOWN BELOW]

[PROGRAMMER: Loop through the following statements in CFIN10 and CFIN10a. CFIN10 is always asked. CFIN10a is only asked if CFIN10 ≤ 3. CFIN10a should be asked immediately after CFIN10 for each item.]

1 Being unable to find child care or being forced to take a child out of child care because they couldn't pay?
2 Falling behind in rent or mortgage payments?
3 Falling behind in gas, electric, or phone bills?
4 Being unable to pay for adequate transportation to get to work or school?
5 Being unable to get medical care because of the cost?
6 Having trouble paying a credit card balance?
7 Having too little money to buy enough food?
8 Being a victim of a crime?
9 Having a problem with alcohol or drug abuse?

USE SHOWCARD #4

1 Always
2 Often
3 Sometimes
4 Rarely
5 Never
7 (VOL) Refused
8 (VOL) Don't know

Base: If the loop iteration of CFIN10 ≤ 3

CFIN10a What ages were you at the time?

["INTERVIEWER: ENTER 0 FOR AGE IF HAPPENED SINCE BIRTH"]

Start age: NUMERIC ENTRY. RANGE = -2-16 [CFIN10a1]
End age: NUMERIC ENTRY. RANGE = -2-16 [CFIN10a2]

7 (VOL) Refused
8 (VOL) Don't know

PROGRAMMER: CHECK CFIN10a1 ≤ CFIN10a2. IF CFIN10a1 > CFIN10a2 DISPLAY "End age should be greater than or equal to start age."

Base: All
CFIN11 From your birth until age 16, did your parents (or guardian) divorce or separate due to financial problems?

DO NOT READ RESPONSES

1 Yes (Go to CFIN11a)
2 No (Go to CFIN12)
7 Refused (Go to CFIN12)
8 Don't know (Go to CFIN12)

Base: If CFIN11 = 1
CFIN11a How old were you at the time? (If necessary: How old were you the first/last time it happened

Age at first time: NUMERIC ENTRY. RANGE = -2-16 [CFIN11a1]
Age at last time: NUMERIC ENTRY. RANGE = -2-16 [CFIN11a2]

-2 (VOL) Don't know
-1 (VOL) Refused

Parents Incarceration[4]

Base: All
CFIN12 Was there ever a time during your childhood that a parent (or guardian) living with you had to serve time in jail or prison?

DO NOT READ RESPONSES

1 Yes (Go to CFIN12a)
2 No (Go to REL1)
7 Refused (Go to REL1)
8 Don't know (Go to REL1)

Base: If CFIN12 = 1
CFIN12a How many years of your childhood was your parent (or guardian) in jail? [IF NEEDED: By childhood, we mean before the age of 18; if more than one sentence, please add the total time, if less than one year, code 0]

Gave response RANGE 0-18
-2 (VOL) Don't know
-1 (VOL) Refused

Base: If CFIN12 = 1
CFIN12b How old were you when your parent (or guardian) went to jail? [IF NEEDED: if more than one sentence, we want the first one]

Start age: NUMERIC ENTRY. RANGE = -2-18 [CFIN12b1]

-2 (VOL) Don't know

[4] From National Survey of Children's Health 2011-12 adapted.

-1 (VOL) Refused

Quality of Early Life Relationships[5]

PROGRAMMER: SKIP ITEM 1 IF CFIN6=6 / SKIP ITEM 2 IF CFIN4=6
Base: IF CFIN4≠6 AND CFIN6≠6

REL1 I am going to read you a series of sentences about your early relationships with your parents. For each of these, please tell me how much you agree with each sentence. There are no right or wrong answers.

[PROGRAMMER: Loop through the following items.]

1 I had a good relationship with my mother before age 18.
2 I had a good relationship with my father before age 18.

[PROGRAMMER: Response options for each item.]

USE SHOWCARD #5

1 Strongly disagree
2 Disagree
3 Neither agree nor disagree
4 Agree
5 Strongly agree
6 (VOL) No parent / Parent deceased
7 (VOL) Refused
8 (VOL) Don't know

[5] HRS.

PROGRAMMER: SKIP REL2 IF CFIN6=6
Base: CFIN6≠6

REL2 I am going to read you a series of sentences. For each of these, please tell me if you would say a lot, some, a little, or not at all. There are no right or wrong answers.

[PROGRAMMER: Loop through the following items.]

1 How much time and attention did your mother give you when you needed it?
2 How much effort did your mother put into watching over you and making sure you had a good upbringing?
3 How much did your mother teach you about life?

[PROGRAMMER: Response options for each item.]

USE SHOWCARD #6

1 A lot
2 Some
3 A little
4 Not at all
6 (VOL) No mother / Mother deceased
7 (VOL) Refused
8 (VOL) Don't know

Base: All

REL3 When you were a child, were there any adults in your life that you felt you could depend on 100% of the time?

DO NOT READ OPTIONS

1 Yes (Go to REL4)
2 No (Go to PAR1)
7 Refused (Go to PAR1)
8 Don't know (Go to PAR1)

Base: If REL3 = 1

REL4 How many of these people were there?

NUMERIC ENTRY. INTEGERS ONLY RANGE = -2, -1, 1 TO 99.

-2 (VOL) Don't know
-1 (VOL) Refused

Base: If REL3 = 1

REL5 Can you please describe how each of these people was related or known to you?

[PROGRAMMER: MULTIPLE RESPONSE]

DO NOT READ RESPONSES, CODE ALL THAT APPLY

1 Mother/Stepmother
2 Father/Stepfather
3 Sibling
4 Aunt/uncle
5 Cousin
6 Grandparent
7 Teacher
8 Family friend
9 Religious leader
10 Coach
11 Other: specify (TEXT ENTRY)
12 Other unrelated person
96 No other mentions
97 Refused
98 Don't Know

Parental Behavior[6]

Base: All

PAR1 I am going to read you a series of sentences. For each of these, please tell me how often each of these happened to you. There are no right or wrong answers. Please think about when you were between the ages of 5 and 12.

[IF NEEDED: Whoever was most like a parent to you.]
[PROGRAMMER: Loop through the following items.]

1 How often would your parent say that you did something that gave him/her pleasure and enjoyment?
2 How often would your parent say that you did something that greatly irritated him/her and got on his/her nerves?
3 How often did your parent read to you?
4 How often did your parent physically punish you as a child, for example by a spanking?
5 How often did your parent praise you as a child, by saying something like "Good for you!" "What a nice thing you did!" "Thank you!" or "That's good going!"
6 How often did your parent tell you about his/her experience, by saying something like, "I saw a pretty bird outside just a little while ago," or "I exercised so hard that I got really tired," or "I was able to give some directions today to somebody that got lost," or "I really like the way the sky looks now."
7 How often did you and your parent talk or play with each other, focusing attention on each other for five minutes or more, without your parent asking or telling you to do anything?
8 How often did your parent tell you to do something, with an irritated or angry tone of voice?
9 How often did you and your parent engage in make-believe play, where you each played the part of a character, and together made up a story to act out with each other?
10 How often did you laugh with your parent?
11 How often did your parent yell or speak to you in a very loud voice, with irritated or angry emotion?

[PROGRAMMER: Response options for each item.]
USE SHOWCARD #7

1 Never
2 Less than once a week
3 About once a week
4 About three or four times a week
5 About once a day
6 Several times each day
7 Many times a day
96 (VOL) No parental figure
97 (VOL) Refused
98 (VOL) Don't know

[6] Rephrased from Parent Practices Survey.

Base: All

PAR2 Thinking about when you were between the ages of 5 and 12…

[IF NEEDED: Whoever was most like a parent to you.]

[PROGRAMMER: Loop through the following items.]

1 What fraction of days did you get three meals, one in the morning, one around noon, and one in the evening?
2 What fraction of days did you get a bath or shower at one particular time, known as your bath time?
3 What fraction of days did you eat all of the following: some meat (or other high protein food), some fruits or vegetables, some milk products, and some bread or grain products?
4 When you and your parent set out to do something fun together, what fraction of the time did it actually turn out to be fun?
5 What fraction of days was your parent too worn out and exhausted to do something fun with you?
6 How often do you think the thought went through your parent's mind that he wished he didn't have to spend so much time with you?
7 Think of all the times that your parent commented to you about your behavior. What fraction were congratulation or approval?
8 Think of all the times that your parent commented to you about your behavior. What fraction were correction or disapproval?

[PROGRAMMER: Response options for each item.]

USE SHOWCARD #8

1 Never
2 Some, but less than a quarter of the time
3 Between a quarter and half the time
4 Not all the time, but more than three quarters of the time
5 All the time
6 (VOL) No parental figure
7 (VOL) Refused
8 (VOL) Don't know

Base: All

PAR3 What fraction of the time did you go to bed at one particular time, known as your bedtime when you were age 9?

[IF NEEDED: Whoever was most like a parent to you.]

1 There was no regular or official bedtime
2 There was an official bedtime, but it was never kept
3 There was an official bedtime and it was kept some, but less than a quarter of the time
4 There was an official bedtime and it was kept between a quarter and half the time
5 There was an official bedtime and it was kept between half and three quarters of the time
6 There was an official bedtime and it was kept not all the time, but more than three quarters of the time
7 There was an official bedtime and it was kept all the time
96 (VOL) No parental figure
97 (VOL) Refused
98 (VOL) Don't know

Base: All

PAR4 Suppose when you were 9, you were handling an object that your parent definitely did not want you to handle. Suppose he or she told you to put the object down, and you defiantly said "No!" Of the following options, which would your parent respond with most of the time?

[IF NEEDED: Whoever was most like a parent to you.]

1 Spank you
2 Send you to your room for half an hour or more
3 Yell at you
4 Repeat the request until you obeyed
5 Ignore you
6 Send you to a room for two to five minutes
7 Send you to a room for five to thirty minutes
8 Show some disapproval in his/her voice and in his/her face, and physically get the object from you, and from then on, if possible, keep the object in a place you couldn't reach
9 (VOL) Something else (Specify: [TEXT ENTRY])
96 (VOL) No parental figure
97 (VOL) Refused
98 (VOL) Don't know

Base: All

PAR5 Did your parent keep you from seeing television shows and movies that had a lot of violence or meanness in them between the ages of 5 and 12?
[IF NEEDED: Whoever was most like a parent to you.]
DO NOT READ RESPONSES

1 Yes
2 No
6 (VOL) No parental figure
7 Refused
8 Don't know

Base: All

PAR6 How often did you see adults or teenagers in your house physically fighting with or hitting or otherwise trying to hurt each other when you were between the ages of 5 and 12? Would you say…

[IF NEEDED: Whoever was most like a parent to you.]

1 Never
2 Less than once a week
3 About once a week
4 About three or four times a week
5 About once a day
6 Several times each day
7 Many times each day
96 (VOL) No parental figure
97 (VOL) Refused
98 (VOL) Don't know

Base: All

PAR7 When your parent gave you a command or order to do something, what fraction of the time did he or she make sure that you did it?

[IF NEEDED: Whoever was most like a parent to you.]

1 Never
2 Some, but less than a quarter of the time
3 Between a quarter and half the time
4 Between half and three quarters of the time
5 Not all the time, but more than three quarters of the time
6 All the time
96 (VOL) No parental figure
97 (VOL) Refused
98 (VOL) Don't know

Base: All

PAR8 Did your parents arrange the objects in your house so that those things that they didn't want you to mess with were not within your reach, so that they didn't have to command you to stay out of them? Would you say…

[IF NEEDED: Whoever was most like a parent to you.]

1 Many things were in reach that a child should leave alone
2 A good number of things were in reach that a child should leave alone
3 A few things were in reach that a child should leave alone
4 Almost no things were in reach that a child should leave alone
5 No things were in reach that a child should leave alone
6 (VOL) No parental figure
7 (VOL) Refused
8 (VOL) Don't know

Base: All

PAR9 Thinking about when you were ages 5 to 12…

[PROGRAMMER: Loop through following items]

1 How often were you able to get your way by having a tantrum?
2 How often did you parent tell you he or she may leave you if you didn't behave better?
3 How often were you punished for crying?
4 How often were you punished for wetting yourself?
5 How often did your parent or someone else tell you that you are bad or not as good as someone else?
6 How often did you see an adult in the house raise his voice in anger at some other adult in the house?
7 How often did you see an adult in the house do something kind, friendly, or very much appreciated by another adult in the house?

[PROGRAMMER: Response options for each item]

USE SHOWCARD #9

1 Never
2 Less than once a week
3 About once a week
4 About three or four times a week
5 About once a day
6 Several times each day
7 Many times each day
96 (VOL) No parental figure
97 (VOL) Refused
98 (VOL) Don't know

Base: All

PAR10 When you asked your parent a question, what fraction of the time did he or she answer it in an enthusiastic and interested way, rather than an irritated way? Would you say your parent…

1 Never answered enthusiastically
2 Answered enthusiastically some, but less than a quarter of the time
3 Between a quarter and half the time
4 Between half and three quarters of the time
5 Not all the time, but more than three quarters of the time
6 Answered enthusiastically all the time
96 (VOL) No parental figure
97 (VOL) Refused
98 (VOL) Don't know

Childhood Health Questions[7]

Show text to all
Consider your health while you were growing up, before you were 16 years old.

Base: All
CHE1 Would you say that your health during that time was excellent, very good, good, fair, or poor?

DO NOT READ RESPONSES

1 Excellent
2 Very good
3 Good
4 Fair
5 Poor
7 Refused
8 Don't know

Base: All
CHE2 When you were growing up, before you were 16 years old, did you miss a month or more of school because of a health problem?

DO NOT READ RESPONSES

1 Yes
2 No
7 Refused
8 Don't know

Base: All
CHE3 Before you were 16 years old, did you have any of the following childhood diseases?

[PROGRAMMER: Loop through following items]

1 Measles
2 Mumps
3 Chicken Pox
4 Asthma
5 Diabetes
6 Epilepsy or seizures

[PROGRAMMER: Response items for each option]

DO NOT READ RESPONSES

1 Yes
2 No
7 Refused
8 Don't know

[7] HRS.

Base: All

CHE4 Before you were 16 years old, did you have difficulty seeing, even with eye glasses or prescription lenses?

DO NOT READ RESPONSES

1 Yes
2 No
7 Refused
8 Don't know

Base: All

CHE5 Did your parents or guardians smoke during your childhood?

DO NOT READ RESPONSES

1 Yes (including some or some of the time)
2 No
7 Refused
8 Don't know

Base: All

CHE6 Before you were 16 years old, did you have…

[PROGRAMMER: Loop through following items]

1 Bronchitis, wheezing, hay fever, shortness of breath or sinus infection or something else that made it difficult to breathe
2 A speech impairment (IF NEEDED: A speech impairment means having difficulty speaking.)
3 Allergies
4 Heart trouble
5 Chronic ear problems or infections (IF NEEDED: Chronic means lasting for a long time)
6 Severe headaches or migraines
7 Stomach problems
8 High blood pressure
9 Depression
10 Drug or alcohol problems
11 Any other emotional or psychological problems

[PROGRAMMER: Response items for each option]

DO NOT READ RESPONSES

1 Yes
2 No
7 Refused
8 Don't know

Base: All

CHE7 Before you were 16 years old, did you have a blow to the head, a head injury or head trauma that was severe enough to require medical attention, or to leave you unconscious or to cause memory loss for a period of time?

DO NOT READ RESPONSES

1 Yes
2 No
7 Refused
8 Don't know

Base: All

CHE8 Before you were 16 years old, were you ever disabled for six months or more because of a health problem? That is, were you unable to do the usual activities of classmates or other children your age?

DO NOT READ RESPONSES

1 Yes (Go to CHE8a)
2 No (Go to CHE9)
7 Refused (Go to CHE9)
8 Don't know (Go to CHE9)

Base: If QCHE_8 = 1

CHE8a What was the cause of that disability?

TEXT ENTRY

-2 (VOL) Don't know
-1 (VOL) Refused

Base: All

CHE9 In grade school or high school, did you ever have a problem in learning the usual lessons, and had to regularly attend special classes, receive special training sessions, or attend a different school for more than six months?

DO NOT READ RESPONSES

1 Yes
2 No
7 Refused
8 Don't know

Base: All

CHE10 Were there any other important or serious health problems that you had before age 16 that you could tell me about?

TEXT ENTRY

-3 (VOL) No such problem
-2 (VOL) Don't know
-1 (VOL) Refused

Childhood Risky Attitudes and Behaviors

Base: All

RATT1 Please think back to when you were 16. I am going to read a list of items that describe feelings or thoughts people sometimes have. Please tell me if at the age of 16, you would have said that the item is true or often true, sometimes true, or not true of you.

[PROGRAMMER: Loop through following items]

1 I had trouble concentrating or paying attention
2 I lied or cheated
3 I teased others a lot
4 I disobeyed my parents
5 I had trouble sitting still
6 I had a hot temper
7 I would rather have been alone than with others
8 I hung around with kids who got into trouble
9 I disobeyed at school
10 I didn't get along with other kids
11 I had trouble getting along with teachers

[PROGRAMMER: Response items for each option]

SHOWCARD #10

1 True
2 Often true
3 Sometimes true
4 Not true
7 (VOL) Refused
8 (VOL) Don't know

Show text to all

Please answer the next few questions based on your behavior as a teenager.

Base: All
RATT2 First, I would like to ask you about smoking habits. As a teenager, did you smoke cigarettes?

DO NOT READ RESPONSES

1 Yes (Go to RATT3)
2 No (Go to RATT5)
7 Refused (Go to RATT5)
8 Don't know (Go to RATT5)

Base: If RATT2 = 1
RATT3 At the age of 16, how many days in a month did you smoke a cigarette?

NUMERIC ENTRY. INTEGERS ONLY. RANGE = -2 TO 31

-2 (VOL) Don't know
-1 (VOL) Refused

Base: If RATT2 = 1
RATT4 When you smoked a cigarette at that time, how many cigarettes did you usually smoke each day? (Note to interviewers: Assume a pack contains 20 cigarettes.)

NUMERIC ENTRY. INTEGERS ONLY. RANGE = -2 TO 99

-2 (VOL) Don't know
-1 (VOL) Refused

Show text to all
Next I would like to ask you some questions about drinking alcoholic beverages, including beer, wine, or liquor.

Base: All
RATT5 As a teenager, did you ever have a drink of an alcoholic beverage? By a drink we mean a can or bottle of beer, a glass of wine, a mixed drink, or a shot of liquor. Do not include childhood sips that you might have had from an older person's drink.

DO NOT READ RESPONSES

1 Yes (Go to RATT6)
2 No (Go to RATT8)
7 Refused (Go to RATT8)
8 Don't know (Go to RATT8)

Base: If RATT5 = 1
RATT6 At the age of 16, on how many days in a month did you have one or more drinks of an alcoholic beverage?

NUMERIC ENTRY. INTEGERS ONLY. RANGE = -2 TO 31

-2 (VOL) Don't know
-1 (VOL) Refused

Base: If RATT5 = 1
RATT7 On the days that you drank alcohol, about how many drinks did you usually have?

NUMERIC ENTRY. INTEGERS ONLY. RANGE = -2 TO 99

-2 (VOL) Don't know
-1 (VOL) Refused

Base: All
RATT8 As a teenager, did you ever use marijuana (that is grass or pot)?

DO NOT READ RESPONSES

1 Yes (Go to RATT9)
2 No (Go to RATT10)
7 Refused (Go to RATT10)
8 Don't know (Go to RATT10)

Base: If RATT8 = 1
RATT9 At the age of 16, on how many days in a month did you use marijuana?

NUMERIC ENTRY. INTEGERS ONLY. RANGE = -2 TO 31

7 (VOL) Refused
8 (VOL) Don't know

Base: All
RATT10 Excluding marijuana and alcohol, as a teenager, did you ever use any other drugs like cocaine or crack or heroin, or any other substance not prescribed for you by a doctor, in order to get high or to achieve an altered state?

DO NOT READ RESPONSES

1 Yes (Go to RATT11)
2 No (Go to text before PFRI1)
7 Refused (Go to text before PFRI1)
8 Don't know (Go to text before PFRI1)

Base: If RATT10 = 1
RATT11 About how many times did you use any of these drugs or other substances in a year?

NUMERIC ENTRY. INTEGERS ONLY. RANGE = -2 TO 365

-2 (VOL) Don't know
-1 (VOL) Refused

Social Networks of Parents[8]

Show text to all
Now I'd like to ask you a few questions about any friendships your parents or guardians had while you were growing up.

Base: All
PFRI1 About how many close friends did your parents have when you were a child? These are people they felt at ease with, could talk to about private matters, or call on for help. Would you say that your parents had no close friends, one or two, three to five, six to ten, or more than ten?

DO NOT READ RESPONSES

1 No close friends (Go to SUC1)
2 1 or 2 (Go to PFRI2)
3 3 to 5 (Go to PFRI2)
4 6 to 10 (Go to PFRI2)
5 More than 10 (Go to PFRI2)
7 Refused (Go to SUC1)
8 Don't know (Go to SUC1)

Base: If PFRI1 = 2,3,4,5
PFRI2 About how many of your parents' close friends…

[PROGRAMMER: LOOP THROUGH FOLLOWING ITEMS]

1 Lived in the same neighborhood as you
2 Graduated from college
3 Worked full-time
4 Were a different race or ethnicity than you

[PROGRAMMER: Response options for each item.]

USE SHOWCARD #11

1 All
2 Most
3 Some
4 A few
5 None
7 (VOL) Refused
8 (VOL) Don't know

[8] Adapted from MTO.

Ideas of Success

SUC1 In a few words, what is your idea of success?

TEXT ENTRY

-2 (VOL) Don't know
-1 (VOL) Refused

Base: All
SUC2 What was your idea of success when you were young?

TEXT ENTRY

-2 (VOL) Don't know
-1 (VOL) Refused

Base: All
SUC3 When you were young, did you believe you would grow up to be successful?

DO NOT READ RESPONSES

1 Yes (Go to SUC4)
2 No (Go to SUC4)
7 Refused (Go to SUC7)
8 Don't know (Go to SUC7)

Base: If SUC3 = 1,2
SUC4 Did you ever (IF SUC3=1: stop / IF SUC3=2: start) believing you would be successful?

DO NOT READ RESPONSES

1 Yes (Go to SUC5)
2 No (Go to SUC7)
7 Refused (Go to SUC7)
8 Don't know (Go to SUC7)

Base: If SUC4 = 1

SUC5 At what age did you (IF SUC3=1: stop / IF SUC3=2: start) believing you would be successful?

NUMERIC ENTRY. RANGE = -2-99

-2 (VOL) Don't know
-1 (VOL) Refused

Base: If SUC4 = 1

SUC6 Why did you (IF SUC3=1: stop / IF SUC3=2: start) believing you would be successful?

TEXT ENTRY

-2 (VOL) Don't know
-1 (VOL) Refused

Base: All

SUC7 Was there a time in your life when help could have made all the difference?

DO NOT READ RESPONSES

1 Yes
2 No (Go to SELF)
7 Refused (Go to SELF)
8 Don't know (Go to SELF)

Base: SUC7 = 1

SUC8 At what age was that time in your life that help could have made all the difference?

NUMERIC ENTRY. RANGE = -2-99

-2 (VOL) Don't know
-1 (VOL) Refused

Base: SUC7 = 1

SUC9 What type of help would have made the difference?

TEXT ENTRY

-2 (VOL) Don't know
-1 (VOL) Refused

Brief Self-Control Scale

Base: All

SELF I am going to read you a series of statements. For each of these, please indicate how much the statement reflects how you typically are on a 1 to 5 scale where 1 means "Not at all" and 5 means "Very much".

[PROGRAMMER: Loop through the following items.]

1 I am good at resisting temptation.
2 I have a hard time breaking bad habits.
3 I am lazy.
4 I say inappropriate things
5 I do certain things that are bad for me, if they are fun.
6 I refuse things that are bad for me.
7 I wish I had more self-discipline.
8 People would say that I have iron self-discipline.
9 Pleasure and fun sometimes keep me from getting work done.
10 I have trouble concentrating.
11 I am able to work effectively toward long-term goals.
12 Sometimes I can't stop myself from doing something, even if I know it is wrong.
13 I often act without thinking through all the alternatives.

[PROGRAMMER: Response options for each item.]

USE SHOWCARD #12

1 1 - Not at all
2 2
3 3
4 4
5 5 - Very much
7 (VOL) Refused
8 (VOL) Don't know

Rotter's Locus of Control Scale

Show text to all

I am going to read you a series of pairs of statements. For each pair, please select the statement that is closer to your opinion. In addition, tell me whether the statement you select is much closer to your opinion or slightly closer. In some cases, you may find that you believe both statements; in other cases you may believe neither one. Even when you feel this way about a pair of statements, select the one statement which is more nearly true in your opinion. Try to consider each pair of statements separately when making your choices; do not be influenced by
your previous choices.

Base: All

LOC1 "What happens to me is my own doing" or "Sometimes I feel that I don't have enough control over the direction my life is taking."

[RANDOMIZE ORDER OF STATEMENTS IN STEM]

1 What happens to me is my own doing.
2 Sometimes I feel that I don't have enough control over the direction my life is taking.
7 (VOL) Refused
8 (VOL) Don't know

LOC1a Would you say this statement is much closer to your opinion or slightly closer?

1 Much closer
2 Slightly closer

Base: All

LOC2 "When I make plans, I am almost certain that I can make them work" or "It is not always wise to plan too far ahead because many things turn out to be a matter of good or bad fortune anyhow."

[RANDOMIZE ORDER OF STATEMENTS IN STEM]

1 When I make plans, I am almost certain that I can make them work.
2 It is not always wise to plan too far ahead because many things turn out to be a matter of good or bad fortune anyhow.
7 (VOL) Refused
8 (VOL) Don't know

LOC2a Would you say this statement is much closer to your opinion or slightly closer?

1 Much closer
2 Slightly closer

Base: All

LOC3 "In my case getting what I want has little or nothing to do with luck" or "Many times we might just as well decide what to do by flipping a coin."

[RANDOMIZE ORDER OF STATEMENTS IN STEM]

1 In my case getting what I want has little or nothing to do with luck.
2 Many times we might just as well decide what to do by flipping a coin.
7 (VOL) Refused
8 (VOL) Don't know

LOC3a Would you say this statement is much closer to your opinion or slightly closer?

1 Much closer
2 Slightly closer

Base: All

LOC4 "Many times I feel that I have little influence over the things that happen to me" or "It is impossible for me to believe that chance or luck plays an important role in my life."

[RANDOMIZE ORDER OF STATEMENTS IN STEM]

1 Many times I feel that I have little influence over the things that happen to me.
2 It is impossible for me to believe that chance or luck plays an important role in my life.
7 (VOL) Refused
8 (VOL) Don't know

LOC4a Would you say this statement is much closer to your opinion or slightly closer?
1 Much closer
2 Slightly close

Dweck Mindset Instrument

Base: All

MIND I am going to read you a series of sentences. For each of these, please indicate how much you agree with the sentence.

[PROGRAMMER: Loop through the following items.]

1 You have a certain amount of intelligence, and you really can't do much to change it.
2 Your intelligence is something about you that you can't change very much.
6 You can learn new things, but you can't really change your basic intelligence.

[PROGRAMMER: Response options for each item.]

USE SHOWCARD #13

1 Strongly Agree
2 Agree
3 Mostly Agree
4 Mostly Disagree
5 Disagree
6 Strongly Disagree
7 (VOL) Refused
8 (VOL) Don't know

CAGE-AID

Read following statement to all

I am going to read you a series of questions. For each of these, please answer Yes or No. When thinking about drug use, include illegal drug use and the use of prescription drug other than prescribed.

[PROGRAMMER: Loop through the following items. AID1 will be asked first. If AID1 = 1, AID2 will be asked. AID2 will always follow AID1 for each iteration of the loop where AID1 = 1.]

1 Have you ever felt that you ought to cut down on your drinking or drug use?
2 Have people annoyed you by criticizing your drinking or drug use?
3 Have you ever felt bad or guilty about your drinking or drug use?
4 Have you ever had a drink or used drugs first thing in the morning to steady your nerves or get rid of a hangover?

Base: All

AID1 [Statement from loop]

DO NOT READ OPTIONS

1 Yes (Ask AID2)
2 No (Go to next loop iteration)
7 Refused
8 Don't know

Base: AID1 = 1

AID2 At what age?
IF MULTIPLE EPISODES, RECORD FIRST AND MOST RECENT EPISODES.

Earliest	NUMERIC ENTRY. INTEGERS ONLY. RANGE = -2-99. [AID2_1]
Most recent	NUMERIC ENTRY. INTEGERS ONLY. RANGE = -2-99. [AID2_2]

[PROGRAMMER: Check that AID2_2 is either missing or AID2_2 ≥ AID2_1.]

-2 (VOL) Don't know
-1 (VOL) Refused

GAD 7

Base: All

GAD I am going to read you a series of problems. For each of these, please indicate how often you have been bothered by it over the **last 2 weeks**.

[PROGRAMMER: Loop through the following items.]

1 Feeling nervous, anxious or on edge
2 Not being able to stop or control worrying
3 Trouble relaxing
4 Being so restless that it is hard to sit still
5 Becoming easily annoyed or irritable
6 Feeling afraid as if something awful might happen

[PROGRAMMER: Response options for each item.]

USE SHOWCARD #14

1 Not at all
2 Several days
3 More than half the days
4 Nearly every day
7 (VOL) Refused
8 (VOL) Don't know

Base: All

ANX Have you ever in your life had an attack of fear or panic when all of a sudden you felt very frightened, anxious, or uneasy?

DO NOT READ RESPONSES

1 Yes
2 No
7 Refused
8 Don't know

Patient Health Questionnaire

Base: All

PHQ1 I am going to read you a series of problems. For each of these, please indicate how often you have been bothered by it over the past two weeks.

[PROGRAMMER: Loop through the following items.]

1 Little interest or pleasure in doing things
2 Feeling down, depressed or hopeless
3 Trouble falling asleep, staying asleep or sleeping too much
4 Feeling tired or having little energy
5 Poor appetite or overeating
6 Feeling bad about yourself – or that you're a failure or have let yourself or your family down
7 Trouble concentrating on things, such as reading the newspaper or watching television
8 Moving or speaking so slowly that other people could have noticed. Or, the opposite – being so fidgety or restless that you have been moving around a lot more than usual
9 Thoughts that you would be better off dead or of hurting yourself in some way.

[PROGRAMMER: Response options for each item]

USE SHOWCARD #15

1 Not at all
2 Several days
3 More than half the days
4 Nearly every day
7 (VOL) Refused
8 (VOL) Don't know

Base: If any of PHQ1_1 to PHQ1_9 between 2 and 4

PHQ2 How difficult have those problems made it for you to do your work, take care of things at home, or get along with other people?

1 Not difficult at all
2 Somewhat difficult
3 Very difficult
4 Extremely difficult
7 (VOL) Refused
8 (VOL) Don't know

Base: All

PHQ3 Have you ever in your life had an episode lasting several days or longer when most of the day you felt sad, empty or depressed?

DO NOT READ RESPONSES

1 Yes (Go to PHQ4)
2 No (Go to PHQ5)
7 Refused (Go to PHQ5)
8 Don't know (Go to PHQ5)

Base: If PHQ3 = 1

PHQ4 At what age?

IF MULTIPLE EPISODES, RECORD FIRST AND MOST RECENT EPISODES.

Earliest NUMERIC ENTRY. INTEGERS ONLY. RANGE = -2-99. [PHQ4_1]
Most recent NUMERIC ENTRY. INTEGERS ONLY. RANGE = -2-99. [PHQ4_2]

-2 (VOL) Don't know
-1 (VOL) Refused

[PROGRAMMER: Check that PHQ4_2 is either missing or PHQ4_2 ≥ PHQ4_1.]

Base: All

PHQ5 Have you ever had an episode lasting several days or longer when you lost interest in most things you usually enjoy like work, hobbies, and personal relationships?

DO NOT READ RESPONSES

1 Yes (Go to PHQ6)
2 No (Go to FAM1)
7 Refused (Go to FAM1)
8 Don't know (Go to FAM1)

Base: If PHQ5 = 1

PHQ6 At what age?

IF MULTIPLE EPISODES, RECORD FIRST AND MOST RECENT EPISODES.

Earliest NUMERIC ENTRY. INTEGERS ONLY. RANGE = -2-99. [PHQ6_1]
Most recent NUMERIC ENTRY. INTEGERS ONLY. RANGE = -2-99. [PHQ6_2]

-2 (VOL) Don't know
-1 (VOL) Refused

[PROGRAMMER: Check that PHQ6_2 is either missing or PHQ6_2 ≥ PHQ6_1.]

Family Environment Scale

Base: All

FAM1 I am going to read you some statements about families. For each of these, please indicate whether they were true of the family you lived with while you were between the ages of 5 and 12. You may feel that some of the statements were true for some family members and false for others. Say True if the statement was true for most members. Say False if the statement was false for most members. If the members are evenly divided, decide what was the stronger overall impression and answer accordingly. Remember, we would like to know what your family seemed like to you. So do not try to figure out how other members would have seen your family.

[PROGRAMMER: Loop through the following items.]

1 We fought a lot in our family
2 Family members attended church, [IF Q1_4=1 OR Q1_8=1 OR Q1_9=1: temple / IF Q1=7: mosque / ELSE: synagogue], or another place of worship or Sunday School fairly often
3 Family members were rarely ordered around
4 Family members rarely became openly angry
5 We didn't say prayers in our family
6 There were very few rules to follow in our family
7 Family members sometimes got so angry they threw things
8 We often talked about the religious meaning of Christmas, [IF Q1=7: Ramadan / IF Q1=8: Vesak / IF Q1=9: Diwali / ELSE: Passover], or other holidays
9 There was one family member who makes most of the decisions
10 Family members hardly ever lost their tempers
11 We didn't believe in heaven or hell
12 There were set ways of doing things at home
13 Family members often criticized each other
14 Family members had strict ideas about what is right or wrong
15 There was a strong emphasis on following rules in our family
16 Family members sometimes hit each other
17 We believed that there were some things you just have to take on faith
18 Everyone had an equal say in family decisions
19 If there was a disagreement in our family, we tried hard to smooth things over and keep the peace
20 In our family each person had different ideas about what was right and wrong
21 We could do whatever we wanted to in our family
22 Family members often tried to one-up or out-do each other
23 The Bible [IF Q1_4=1 OR Q1_7=1 OR Q1_8=1 OR Q1_9=1: or another holy book like] [IF Q1_4=1: the Book of Mormon] [IF Q1_7=1: the Koran] [IF Q1_8=1: the Sutras] [IF Q1_9=1: the Vedas] was a very important book in our home
24 Rules were pretty inflexible in our household
25 In our family, we believed you don't ever get anywhere by raising your voice
26 Family members believed that if you sinned, you would be punished
27 You couldn't get away with much in our family

[PROGRAMMER: Response options for items]

1 True
2 False
7 (VOL) Refused
8 (VOL) Don't know

Base: All

FAM2 When you answered these questions, who did you have in mind?

[PROGRAMMER: MULTIPLE RESPONSE]

CODE ALL THAT APPLY

1 Your brothers and sisters
2 Your parent
3 Your roommate
4 Other family members
5 Your spouse or partner
6 Your child or children, or
7 Someone else?
96 (VOL) No other mentions
97 (VOL) Refused
98 (VOL) Don't know

50-item IPIP version of the Big Five Markers

Base: All

PIP I am going to read you a series of statements about yourself. For each of these, please indicate whether it is 1. Very Inaccurate, 2. Moderately Inaccurate, 3. Neither Accurate Nor Inaccurate, 4. Moderately Accurate, or 5. Very Accurate as a description of you. Describe yourself as you generally are now, not as you wish to be in the future. Describe yourself as you honestly see yourself, in relation to other people you know of the same sex as you are, and roughly your same age.

[INTERVIEWER: IF RESPONDENT HAS TROUBLE WITH THE ANSWER CHOICES, PLEASE ASSIST THEM]

[PROGRAMMER: Loop through the following items.]

1 I am the life of the party.
2 I feel little concern for others.
3 I am always prepared.
4 I get stressed out easily.
5 I have a rich vocabulary.
6 I don't talk a lot.
7 I am interested in people.
8 I leave my belongings around.
9 I am relaxed most of the time.
10 I have difficulty understanding abstract ideas.
11 I feel comfortable around people.
12 I insult people.
13 I pay attention to details.

14 I worry about things.
15 I have a vivid imagination.
16 I keep in the background.
17 I sympathize with others' feelings.
18 I make a mess of things.
19 I seldom feel blue.
20 I am not interested in abstract ideas.
21 I start conversations.
22 I am not interested in other people's problems.
23 I get chores done right away.
24 I am easily disturbed.
25 I have excellent ideas.
26 I have little to say.
27 I have a soft heart.
28 I often forget to put things back in their proper place.
29 I get upset easily.
30 I do not have a good imagination.
31 I talk to a lot of different people at parties.
32 I am not really interested in others.
33 I like order.
34 I change my mood a lot.
35 I am quick to understand things.
36 I don't like to draw attention to myself.
37 I take time out for others.
38 I ignore my duties.
39 I have frequent mood swings.
40 I use difficult words.
41 I don't mind being the center of attention.
42 I feel others' emotions.
43 I follow a schedule.
44 I get irritated easily.
45 I spend time reflecting on things.
46 I am quiet around strangers.
47 I make people feel at ease.
48 I am exacting in my work.
49 I often feel blue.
50 I am full of ideas.

[PROGRAMMER: Response options for each item.]

USE SHOWCARD #16

1 Very Inaccurate
2 Moderately Inaccurate
3 Neither Accurate Nor Inaccurate
4 Moderately Accurate
5 Very Accurate
7 (VOL) Refused
8 (VOL) Don't know

Short Grit Scale

Base: All

GRIT I am going to read you a number of statements that may or may not apply to you. For the most accurate score, when responding, think of how you compare to most people – not just the people you know well, but most people in the world. There are no right or wrong answers, so just answer honestly!

[PROGRAMMER: Loop through the following items.]

1 New ideas and projects sometimes distract me from previous ones.
2 Setbacks don't discourage me.
3 I have been obsessed with a certain idea or project for a short time but later lost interest.
4 I am a hard worker.
5 I often set a goal but later choose to pursue a different one.
6 I have difficulty maintaining my focus on projects that take more than a few months to complete.
7 I finish whatever I begin.
8 I am diligent

[PROGRAMMER: Response options for each item.]

USE SHOWCARD #17

1 Very much like me
2 Mostly like me
3 Somewhat like me
4 Not much like me
5 Not like me at all
7 (VOL) Refused
8 (VOL) Don't know

Rosenberg Self-Esteem Scale

Base: All

RSE I am going to read you a series of statements dealing with your general feelings about yourself. For each of these, please indicate how strongly you agree or disagree.

[PROGRAMMER: Loop through the following items.]

1 On the whole, I am satisfied with myself.
2 At times, I think I am no good at all.
3 I feel that I have a number of good qualities.
4 I am able to do things as well as most other people.
5 I feel I do not have much to be proud of.
6 I certainly feel useless at times.
7 I feel that I'm a person of worth, at least on an equal plane with others.
8 I wish I could have more respect for myself.
9 All in all, I am inclined to feel that I am a failure.
10 I take a positive attitude toward myself.

[PROGRAMMER: Response options for each item.]

1 Strongly Agree
2 Agree
3 Disagree
4 Strongly Disagree
7 (VOL) Refused
8 (VOL) Don't know

Adverse Childhood Experience Questionnaire

Base: All

ACE We are almost at the end of the survey. The questions I am about to read could be distressing to individuals with a history of post-traumatic stress disorder, or survivors of abuse and violence. You may find these questions upsetting. Please feel free to ask to skip these questions or discontinue the interview without penalty. You will not need to explain why you do not wish to answer or stop the interview. Remember that I also gave you a list of local support organizations at the beginning of the interview. I am going to read you a series of questions about your childhood. For each of these, please answer yes or no.

[PROGRAMMER: Loop through the following items.]

1 Did a parent or other adult in the household **often** swear at you, insult you, put you down, or humiliate you **or** act in a way that made you afraid that you might be physically hurt?
2 Did a parent or adult in the household **often** push, grab, slap, or throw something at you **or ever** hit you so hard that you had marks or were injured?
3 Did an adult or person at least 5 years older than you **ever** touch or fondle you or have you touch their body in a sexual way **or** try to actually have oral, anal, or vaginal sex with you when you were a minor? (IF NEEDED: A minor is someone under the age of 18.)
4 Did you **often** feel that no one in your family loved you or thought you were important or special **or** your family didn't look out for each other, feel close to each other, or support each other?
5 Did you **often** feel that you didn't have enough to eat, had to wear dirty clothes, and had no one to protect you **or** your parents were too drunk or high to take care of you or take you to the doctor if you needed it?
6 Were your parents **ever** separated or divorced?
7 Was your mother or stepmother **often** pushed, grabbed, slapped, or had something thrown at her **or sometimes or often** kicked, bitten, hit with a fist, or hit with something hard **or ever** repeatedly hit over at least a few minutes or threatened with a gun or knife?
8 Did you live with anyone who was a problem drinker or alcoholic or who used street drugs?
9 Was a household member depressed or mentally ill or did a household member attempt suicide?
10 Did a household member go to prison?

[PROGRAMMER: Response options for each item.]

DO NOT READ RESPONSES

1 Yes
2 No
7 Refused
8 Don’t know

Closing

Base: All

PAYINFO Those are all the questions I have. Thanks so much. In order to document that you received your $150 VISA gift card for participating, I just need to record your full name:

FULL NAME

INTERVIEWER: RECORD GIFT CARD NUMBER HERE:

Thank you. My Supervisor or another staff member may call you to check that I talked to you today. It will only take a few minutes and would be a big help if you speak with them.

Have a great day!

E Literature Review of the Correlates of Intergenerational Mobility

Authors	Data	Sample	Correlates
McKernan and Ratcliffe (2005)	Monthly, longitudinal data from the 1988, 1990, and 1996 panels of the Survey of Income and Program Participation (SIPP) and monthly unemployment rates from the U.S. Department of Labor (2001) with quarterly real gross domestic product from the U.S. Department of Commerce (2001).	Sample size for each panel ranges from 14,000 to 36,700 households and data is collected for the preceding four months in each interview of the SIPP participant.	Education
Haskins (2008)	Panel Study of Income Dynamics (PSID) that tracks the mobility of adult children by comparing their income at roughly age 40 with that of their parents at about the same age.	All participants of the PSID survey where parents' income was averaged over the period 1967-1971 and adult children's incomes were averaged over selected years between 1995 and 2002.	Education
Baum, Ma and Payea (2013)	Panel Study of Income Dynamics (PSID)	PSID participants where parents' incomes have been taken from 1967 through 1971. Income for children is computed for the years 2000, 2002, 2004, 2006 and 2008.	Education
Aldaz-Carroll and Moran (2001)	Panel data from the Lima Panel survey conducted in Lima, Peru in 1985 and 1994. The data contains educational attainment, family background, health, information on domestic violence and poverty status.	294 individuals that are participants in the Lima Panel and are between the ages of 16 and 26 in 1994.	Education, parent's education, gender, single mother and urban/rural residence.

Metzler et al. (2017)	2010 Behavioral Risk Factor Surveillance System (BRFSS) that collects data on Adverse Childhood Experiences, education attainment, unemployment and poverty status.	Data from 27,834 noninstitutionalized adults surveyed from 10 states and the District of Columbia that used the adverse childhood experiences (ACE) module in the 2010 Behavioral Risk Factor Surveillance System.	Adverse Childhood Experiences
Duncan and Rodgers (1988)	1968-1982 waves of the Panel Study of Income Dynamics that contains data on family background, educational attainment and income.	1075 children who were under the age of 4 in 1968 who had not left home before the end of a 15-year period and for whom a full 15 years of family data were available.	Family events
Sharkey and Torrats-Espinosa (2015)	Equality of Opportunity project that contains data on intergenerational economic mobility, crime and community zone demographics.	287 urban commuting zones for which authors have non-missing data on crime and economic mobility.	Violent crime
Wastern and Pettit (2009)	National Longitudinal Survey of Youth 1979 (NLSY 1979)		Incarceration rates
Freeman (1985)	1979-80 National Bureau of Economic Research-Mathematica survey of Inner City youth and the 1979-81 National Longitudinal Survey of Young Men with data on church going habits, time allocations, school, work activity, frequency of socially deviant activity, family background and poverty status.		Church-going

Keels et al. (2005)	Information on the Gautreax program participants were provided by the Leadership Council. This information contained mother's age, AFDC recipiency status and number of children. Addresses were taken from a credit reporting service and the Illinois Department of Human Services Integrated Client Database Records. Neighborhood characteristic were calculated from the 1980 and 1990 U.S. censuses and crime data for neighborhoods were taken from FBI's UCR records.	1506 randomly chosen families who moved as part of the Gautreaux program prior to 1990.	Neighborhood
Sanbonmatsu et al. (2012)	Interviews of adults from "Moving to Opportunity" (MTO) households 10 to 15 years after families were randomized into the program. Data includes adult's health and economic circumstances, physical measurements and blood samples.	3273 adults from MTO households.	Neighborhood
Chetty et al. (2014)	Incomes from the Internal Revenue Service Databank between 1996 and 2012 and neighborhood demographocs from 1990 Census.	6.3 million U.S citizens in 1980-1981 birth cohorts between the ages of 30 and 32 in 2012.	Neighborhood
Chetty, Hendren and Katz (2016)	"Moving to Opportunity" (MTO) Data linked to federal income tax returns	4604 families in 5 large US cities that were randomized as part of the MTO experiment.	Neighborhood

Chetty and Hendren (2017)	Census data (2000, 2010, ACS) covering U.S population linked to federal income tax returns from 1989-2015.	Children in 1978-83 birth cohorts who were born in the U.S or are authorized immigrants who came to the U.S in childhood. Analysis sample covered 20 million children.	Neighborhood

F Literature Review of the Causal Impacts of Important Variables

Correlate	Authors	Data	Sample	Causal Impact
Early Childhood Education	Belfield et al. (2006), Heckman et al. (2010), Heckman, Pinto and Savelyev (2013)	Numerous measures ranging from economic, criminal, and educational outcomes as well as cognition and personality between the ages of 3 and 40.	123 HighScope Perry Preschool Program participants who mostly consist of low-IQ, disadvantaged African-American children living in Ypsilanti, Michigan.	Monthly income at age 27 increases by $867 (Heckman, Pinto and Savelyev, 2013).
Early Childhood Education	Barnett and Masse (2007), Elango et al. (2015)	Family background characteristics at study entry, school assessment test scores, college enrollment, crime, personality, health, behavior and earnings.	104 study participants and their families who were sampled to be economically disadvantaged were randomly assigned to the Abecedarian program.	The program effect on the gross earnings of future generations was estimated at $5700 (Barnett and Masse, 2007). Yearly labor incomes (in 2014 USD) increased by $ 3,578 for females and $ 17,214 for males. (Elango et al., 2015).

Early Child-hood Educa-tion	Garces, Thomas and Currie (2002) and Elango et al. (2015)	Panel Study of Income Dynamics which records participation in Head Start and have long term follow up data. Data includes cognitive traits, high school completion, college attendance, crime, health behaviors and earnings.	Almost 4000 adults aged 18 and older in 1995 who were PSID respondents.	Earnings between the ages of 23 and 40 increased by 5.1 percentage points.
Education	Card (1999)		Survey of literature on the causal relationship between education and earnings. Selected studies use institutional aspects of the education system to form insytrumental variables; studies of earnings and schooling of twins; and studies that explicitly model sources of heterogeneity in the returns to education.	An additional year of schooling can increase wages by 2-11%.

Education	Heckman, Humphries and Veramendi (2016)	National Longitudinal Survey of Youth 1979 covering schooling and earnings variables.	Sample of males extracted from NLSY 1979.	Completing an additional year in high school increases wages by 9.4 percentage points. Completing an additional year for college enrollees increases wages by 13.4 percentage points. Completing an additional year for college graduates increases wages by 11.4 percentage points.
Big 5 Personality Traits	Nyhus and Pons (2005)	CentER Saving Survey (CSS) which provides individual labor market details as well as responses to the Five Factor Personality Inventory (FFPI)	888 workers aged 16-65 who were part of the respondent pool for CentER Saving Survey in the Netherlands	Agreableness has a significant and negative effect in the male sample while emotional stability positively affects the wage setting for women.
Big 5 Personality Traits	Heckman, Pinto and Savelyev (2013)	Numerous measures ranging from economic, criminal, and educational outcomes as well as cognition and personality between the ages of 3 and 40.	123 HighScope Perry Preschool Program participants who mostly consist of low-IQ, disadvantaged African-American children living in Ypsilanti, Michigan.	The Perry Pre School treatment increased monthly income at age 27 by $867. 20 percent of this treament effect is explained by early improvements of personality traits.

Self Esteem and Locus of Control	Heckman, Stixrud and Urzua (2006)	National Longitudinal Survey of Youth 1979 which follows a nationally representative sample of 14-22 year olds between 1979 and 2014. The dataset contains information on wage, schooling, employment, cognitive and non-cognitive measures.	All participants of NLSY 1979.	A 1 standard deviation increase in non-cognitive ability increases hourly wages by 11.2%.
Self control	Moffitt et al. (2011)	Dunedin Multidisciplinary Health and Development Study and Environmental Risk Longitudinal Study (E-Risk). The dataset contains measures of self control, health outcomes, school performance, behaviors and earnings.	The Dunedin Study tracks a complete birth cohort of 1037 children from birth to the age of 32 years. The E-Risk Study tracks 500 sibling pairs since their birth in 1994 and 1995 to their adulthood.	

Child Abuse	Currie and Widom (2010)	Using a prospective cohort design, court substantiated cases of childhood physical and sexual abuse and neglect during 1961 - 1971 were matched with nonabused and nonneglected children in one Midwestern metropolitan county area and followed into adulthood (mean age 41). Data contains demographics, family background, education and other cognitive tests, psychiatric tests, and economic status and productivity in 2003-2004.	807 individuals who were part of the prospective cohort design and had non-missing outcome measures in 2003- 2004.	Child maltreatment reduces earnings in 2003-2004 by aprroximately $5000.
Youth Criminal Behavior	Allgood, Mustard and Warren, Jr. (2007)	National Longitudinal Survey of Youth 1979 that contains detailed information on criminal background, socioeconomic and demographic characeristics from 1980 and earnings from 1983 and 1989.	439 sibling pairs from NLSY 1979 sample.	A criminal charge when young reduces 1989 earnings by 22% while a conviction reduces 1989 earnings by 36%.