Pollution and the Intergenerational Transmission of Human Capital: Evidence from the 1970 Clean Air Act.*

Jonathan Colmer
University of Virginia

John Voorheis
U.S. Census Bureau

Preliminary and Incomplete – Please Do Not Cite or Circulate Without Permission

Abstract

Using a newly constructed dataset linking administrative, survey and decennial Census data, we evaluate the intergenerational effects of early life pollution exposure. Exploiting variation in particulate matter, which sharply dropped following the enactment of the 1970 Clean Air Act Amendments, we find that the children of those affected by additional improvements in air quality are more likely to attend college. Furthermore, we find no differential effect between the adopted and biological children of affected parents, and find suggestive evidence that parents who experienced large declines in pollution exposure are more likely to engage in child enrichment activities. This suggests that the transmission mechanism arises through parental investments and resources, rather than genetic channels.

*This Version: December 2018. Correspondence: j.colmer@virginia.edu. Affiliations: Colmer – Department of Economics, University of Virginia, USA; Voorheis – U.S. Census Bureau, USA. We thank Anna Aizer, Doug Almond, Max Auffhammer, Leora Friedberg, Bill Johnson, Nicolai Kuminoff, Maya Rossin-Slater, Chris Ruhm, Nicholas Sanders, Jay Shimshack, Bill Shobe, Sandip Sukhtankar, Reed Walker, James Sallee, Lucas Davis and Catherine Wolfram for many helpful thoughts, comments, and discussions. We are also grateful to seminar participants at Arizona State University, UC Berkeley, Columbia University, the Heartland Environmental & Resource Workshop at Illinois, the University of Oregon, the University of Virginia and the U.S. Census Bureau for helpful comments and suggestions. All errors and omissions remain our own. This paper is released to inform interested parties of research and to encourage discussion. The views expressed are those of the authors and not necessarily those of the U.S. Census Bureau. All results have been reviewed to ensure that no confidential information is disclosed. The statistical summaries reported in this paper have been cleared by the Census Bureau's Disclosure Review Board, release authorization numbers CBDRB-FY18-232, CBDRB-FY18-248, CBDRB-FY18-287, CBDRB-FY19-099, CBDRB-FY19-136, CBDRB-FY19-261.

1 Introduction

Economists and social scientists have long been interested in the intergenerational transmission of parental endowments. However, little is known about the causal effects of changes to the economic prospects of one generation on the outcomes of the second generation, and even less is known about the drivers underlying intergenerational mobility.

Growing evidence, in both epidemiology and economics, suggests that in-utero exposure to air pollution, and other environmental factors, can play a substantial role shaping endowments at birth, resulting in persistent long-term effects on health and welfare. However, evidence on the intergenerational transmission of environmentally-induced shocks to endowments is far more limited.[1] Consequently, within-generation estimates of the dose-response function between environmental conditions and well-being may substantially underestimate the total welfare impact of environmental toxins. In this paper we present some of the first evidence on the intergenerational transmission of parental endowments, shaped by early-life environmental exposure.

We exploit the introduction of the 1970 Clean Air Act Amendments (CAAA), which imposed county-level restrictions on the maximum-allowable concentrations of total suspended particulates (TSP). Any counties that exceeded the new regulatory ceiling (non-attainment counties) were forced to reduce their TSP concentrations below the ceiling, while counties that had air pollution levels that were already below the regulatory ceiling (attainment counties) were not legally required to reduce their TSP emissions. This resulted in substantial reductions in ambient air pollution levels in hundreds of counties across the United States. We then exploit variation in exposure to these changes by comparing outcomes for the children of cohorts that were born just before these changes went into effect to the children of cohorts that were born just after these large changes in air pollution, isolating any additional impacts of exposure to clean air in very early childhood relative to exposure at slightly older

[1] A notable exception being Black et al. (2018) who examine the intergenerational consequences of nuclear testing in Norway between fathers and sons.

ages.

It is important to note that while the question under study is novel, the quasi-experimental design described is well established. The 1970 CAAA variation has been previously used to study the effects of air pollution on contemporaneous outcomes such as infant mortality (Chay and Greenstone, 2003a) and fetal mortality (Sanders and Stoecker, 2015), as well as later-life outcomes such as adult mortality (Chay et al., 2003), educational attainment (Voorheis, 2017a) and adult earnings (Isen et al., 2017). Consequently, we follow the original research design, implemented by (Chay and Greenstone, 2003a), as closely as possible to discipline our analysis.

We combine this quasi-experimental design with a newly constructed linked dataset, combining administrative and survey data from the U.S. Census Bureau. This linked dataset allows us to identify the exact date and location of birth for the universe of children born in the 1960s and 1970s. We then identify hundreds of millions of parent-child links, allowing us to identify the effects of in-utero exposure to ambient air pollution on the children of those that were in-utero exposed – the second generation effects.

We begin by first examining how the large reductions in pollution exposure induced by 1970 CAAA affected the later life outcomes for individuals born between 1960-1980 – the first generation effects. In terms of economic outcomes, we find that lower in-utero exposure to particulate matter is associated with significant increases in later life earnings. A 10 $\mu g/m^3$ reduction in gestational TSP exposure is associated with a 0.9 percent increase in annual earnings, \$407 on average.[2] Assuming that this effect is constant over the life cycle, and that earnings are discounted at a real rate of 3 percent back to age zero (5 percent discount rate + 2 percent wage growth), the lifetime earnings effect of a 10 $\mu g/m^3$ increase in TSP is \$4,066.11 per person. Furthermore, if we assume a linear dose response function and apply the observed changes in TSP since 1971, the aggregate increase in earnings associated with reduced in-utero TSP exposure come to \$1.978 trillion.

[2]Our findings are similar in magnitude to the estimates provided by Isen et al. (2017), using data for 26 states from the LEHD survey.

In addition to looking at economic outcomes, we also explore changes to family structure, which may be an important consideration in driving the intergenerational transmission of pollution exposure. We find that higher pollution exposure is associated with a very small increase in the likelihood of an affected individual being divorced, but does not affect other family structure outcomes such as the likelihood of getting married. Furthermore, we do not find any effect on the likelihood of having children, the number of children, the likelihood of teen pregnancies, or the timing of children more generally. Understanding the effects of exposure on fertility is important to the degree that it affects selection into the second generation sample. This consideration is largely understudied in the literature. However, our findings suggest that, at least in this context, selection into the second-generation sample does not appear to be a first-order concern.

Next, we explore how the large reduction in pollution exposure experienced by parents born between 1960-1980 affects the human capital acquisition of their children – the second generation effects. Our focus on human capital is largely motivated by timing – very few individuals over the age of 22 at the time of their ACS response have parents born after 1971, making the estimation of parental in-utero pollution exposure on children's wages or labor force participation empirically challenging. However, there are human capital measures available in the ACS which are prevalent among people under 22, who are much more likely to have "treated" parents.

We find that the children whose parents experienced higher in-utero exposure to particulate matter are less likely to attend college highlighting the intergenerational transmission of in-utero pollution exposure. However, there is no difference in the likelihood of dropping out of high school or being held back a grade, suggesting that the mechanism through which transmission occurs is not relevant to human capital investments earlier in life.[3]

Given the significant differences in lifetime earnings for those with a college degree com-

[3]A caveat to this is that being held back a year and dropping out of high school drop-out are relatively extreme and extensive margin events. There may be intensive margin effects on human capital in the early years that are not captured here.

pared to those without we expect that the economic consequences of the estimated college enrollment effect are substantial. Using a college wage premium of $25,000 p.a. (estimated using Mincerian wage regressions), combined with a graduation rate of 50 percent to convert college attendance to college completion (NCES), and a 3 percent real discount rate we estimate a $10\mu g/m^3$ reduction in parental gestational TSP exposure amounts to a $703 increase in discounted lifetime earnings at (first-gen) age zero.[4] Combining the number of births since 1971, the probability that an individual has a child (0.61), and the average number of children had by each individual (1.4), the aggregate benefits of observed TSP reductions since 1971 to the second generation amount to $292 billion.

Combining the first and second generation estimates, the aggregate earnings impact of in-utero pollution exposure is 17-58 percent higher than our estimates based on a within-generation analysis, suggesting that a large part of the pollution exposure effect on earnings is likely to be passed on to the second generation.

Having estimated the persistence of exposure across generations we also seek to understand the mechanism that underlies these effects. Broadly speaking there are two channels. The first channel is genetic. If increased in-utero exposure to pollution results in epigenetic changes – permanent changes to gene expression – for the first generation then these changes may be passed on to the second generation, directly affecting human capital and potentially reducing the returns to college.[5] The second mechanism is economic. Given the effects of in-utero exposure on parental earnings and the potential for changes in family structure, through divorce, parental resources and investments may affect the opportunity of the second-generation to attend college.

We begin by exploring the relevance of the genetic pathway. Using information on the adopted status of children, i.e., whether children are adopted or biological we examine the

[4]Discounting back to age zero for the second generation delivers a lifetime earnings estimate of $2,379 (58 percent of the first generation effect.

[5]Note that as a consequence of our research design there should be no genetic differences, on average, between individuals born in non-attainment and attainment counties and so the only genetic mechanism that can arise is epigenetic.

differential effects of parental in-utero TSP exposure on college attendance for adopted children, compared to biological children. We find no differential effects, suggesting that the transmission mechanism across generations arises through parental resources rather than genetic channels. Importantly, we examine the effects of in-utero exposure on the propensity to adopt for the first-generation, evaluating the degree to which there may be selection into the second-generation sample along this margin. We find no effects of in-utero pollution exposure on the propensity to adopt.

These results suggest that parental resources and investments underlie the transmission of human capital across generations in this context. Current work seeks to understand as precisely the relevant channels through which this effect arises.

Our findings contribute to several literatures. First, we contribute to an established literature documenting the importance of environmental factors, as opposed to genetic factors, in determining human capital endowments at birth (Chay and Greenstone, 2003a,b; Almond, 2006; Black et al., 2007; Currie et al., 2009; Fertig and Watson, 2009; Kelly, 2011; Almond et al., 2010; Isen et al., 2017; Black et al., 2018). However, these papers have largely focussed on educational and labor market outcomes for the first-generation only. By contrast, we extend our analysis to also explore effects on family structure for the first generation, an area for which causal evidence is more limited (Gruber, 2004), as well as exploring the persistence of environmentally-driven endowments at birth across generations.

Second, we contribute to an emerging literature on the distributional consequences of environmental change (Hsiang et al., 2018). To date much of this work has focused on "inequality at birth" (Currie, 2011). By contrast, we seek to understand the degree to which inequality may persist over time as well as across individuals within a generation. Furthermore, if improvements in air quality are complementary to parental inputs through human capital then these improvements may further reduce inequality over time, through a crowding in of human capital.

Finally, we contribute to a broad literature seeking to understand the intergenerational transmission of human capital (Becker and Tomes, 1979; Solon, 1992; Black et al., 2005, 2007; Black and Devereux, 2011; Chetty et al., 2014; Chetty and Hendren, 2018a,b; Chetty et al., 2018). We build on this literature in several ways. First, traditionally this literature has focused on measuring and documenting the persistence of wealth, income, or human capital across generations. By contrast, we contribute to a nascent literature focused on understanding the causal effect of shocks to parental endowments on later generations (Black et al., 2018; Barr and Gibbs, 2017; East et al., 2017; Akresh et al., 2018). Second, we provide new insights into the mechanisms that underlie the transmission of parental endowments, finding that the transmission of human capital appears to arise predominantly through parental resources and investments, rather than through genetic channels.[6]

Collectively, our findings highlight the relevance of intergenerational spillovers in response to environmental change with important implications for understanding the distributional consequences of environmental change, as well as for our understanding of how policy – in this context environmental policy – affects intergenerational mobility.

The remainder of the paper is structured as follows. Section 2 describes the data sources as well as the process through which identify parent-child links.[7] Section 3 outlines the various econometric models used, and section 4 discussed the results provided by those models. Finally, section 5 discusses the implications of our findings and concludes.

2 Data

To study the intergenerational effects of the Clean Air Act, it is necessary to locate parents at birth (around the enactment of 1970 CAAA), infer their exposure to ambient air pollution, link these parents to their children, and measure outcomes for both parents and children. No single dataset has all of these features, and so our analysis requires linking

[6]This is not to say that the transmission of human capital through epigenetic channels is not in existence, but that this does not explain the effects we find on college attendance. There may be latent health effects that we do not capture at this stage in the second generations life cycle.

[7]A more complete discussion can be found in a supplementary data Appendix.

multiple datasets from survey, decennial Census and administrative records sources. This linkage is done using unique anonymous personal identifiers called Protected Identification Keys (PIKs). PIKs, which can be thought of as a "scrambled" Social Security Number, are assigned to datasets using a probabilistic matching algorithm which links personally identifiable information (name, date of birth, Social Security Number, etc.) to a reference file of people in the United States.[8]

2.1 Parent-Child Links

We begin by assembling a database of a majority of the parent-child links that can be evaluated using survey, decennial Census and administrative data sources available in the Census Bureau's data linkage infrastructure. We identify links in two main datasets: the full count (short form) decennial Census from 2000 and 2010, and the American Community Survey (ACS) from 2005–2015. The set of links we are able to identify is not, we should stress, the full population of links. We will miss two main sets of parent-child linkages: parent-child linkages in households which formed and dissolved between decennial Censuses (who were not ACS respondents), and parent-child links in which either the parent or child cannot be assigned a PIK.

The decennial Census and ACS data both contain detailed information on relationships within household, with one important limitation – the Census/ACS relationship question asks for information only on the relationship between an individual and the head of household. For our purposes, this means that we can identify parent-child links for the head of household parent with certainty. We additionally identify probable parent-child links between the head of household's married or unmarried partner and the head of household's children. For head of household-child links, we have additional information about the type of link – specifically whether a child is natural born, adopted or a step-child.

[8]For more on the process of PIK assignment see Wagner and Layne (2014).

2.2 Exposure

To analyze the intergenerational effects of pollution exposure, we need to be able to infer the level of ambient air pollution and the changes in Environmental Protection Agency (EPA) policy – designation of nonattainment of air quality standards – that parents were exposed to at birth. We do this in three steps. First, we link the set of unique parents identified above to the Census Numident to obtain date and place of birth. We then obtain monitor-level daily pollution measures from the EPA, which aggregate to the county level, and link these county-level measures to the parents' place of birth. Finally, we simulate these nonattainment designations for counties with EPA monitors active in 1969 (before 1970 CAAA), as in Isen et al. (2017).

The Census Numident is a person-level administrative records file derived from the Social Security's Numident, which contains all individuals who ever apply for a Social Security Number. Importantly, the Numident contains information on individuals' exact date of birth, and place of birth. As the place of birth information is not standardized, we assign county of birth information to individuals using the crosswalk used by Isen et al. (2017) and the probabilistic matching approach used in Voorheis (2017b). We identify county of birth using this approach for both first and second generation individuals.

With information about the place of parents' birth in hand, we infer the level of pollution exposure experienced by these individuals based on the average exposure within their county of birth. To gather this pollution exposure information, we rely on monitor data from the EPA, which we retrieve using a public facing API[9]. Our pollutant of interest is particulate matter. For the relevant period of time (around 1970), the primary regulated pollutant was total suspended particulates (TSP), defined as particulate matter with a density of less than 50 microns, measured in units $\mu g/m^3$. We thus retrieve all TSP monitor observations between 1960–1980.

The TSP standard was set based on a 24-hour sampling, and hence the monitor-level data

[9]See https://aqs.epa.gov/aqsweb/documents/data_mart_welcome.html for more details.

is provided on a daily basis. Our baseline approach for aggregating these daily monitor-level observations is as follows. For each county-day, we calculate the average TSP concentration across all active monitors in that day, which we take as the average exposure to TSP in that county on that day. We then calculate county-level moving average exposure to TSP for each unique birthday between 1960 and 1980 for two periods of interest: the nine months before birth (in utero exposure) and the year after birth (infant exposure).

Finally, our empirical strategy requires information on which counties were designated as in nonattainment of the ambient air quality standards in the 1970 CAAA by the EPA. Although the EPA makes nonattainment designations publicly available starting in 1991, and researchers have reconstructed nonattainment designations back to 1980, there appear to be no existing records on which counties were in nonattainment in 1972, the first year in which the 1970 CAAA was in effect. However, the TSP air quality standards are known, and as noted above, we have monitor-level data on TSP concentrations in the years before the 1970 CAAA was in effect. Thus it is possible to reconstruct which counties would have been in non-attainment.

Nonattainment of the primary air quality standard for TSP set in 1970 CAAA occurs if either a) a county's annual average (geometric mean) TSP concentration is above 75 $\mu g/m^3$, or b) the second highest daily TSP concentration is above 260 $\mu g/m^3$. We use the monitor-level observations discussed above to calculate the geometric mean and second highest daily TSP concentration for all counties with at least on monitor in 1970. This allows us to categorize 258 counties as "nonattainment" counties, and 319 counties as "attainment" counties.[10]

[10] Consequently, we have to restrict our analysis to first generation individuals born in these 577 counties and second generation individuals born to individuals born in these counties, as the pollution exposure of individuals born in other counties was unmeasured during this time period. Note however that these 577 counties contained about two thirds of the US population in 1969 and cover all 50 states.

2.3 Outcomes

Finally, we require information on outcomes for parents and children, as well as other information on observables, such as socio-demographic characteristics. We measure these outcomes using the ACS, which contains detailed information on the family structure, human capital and labor market outcomes we are interested in. Note that since the ACS is a nationally representative survey of a (very large) sample of households, we observe outcomes from only a fraction of the parents and children identified above.

The ACS contains information on marital status, from which we we define variables for being married or divorced at survey response. The ACS microdata contains information about fertility which we use directly: presence of own children (asked to all women of child bearing age) and number of own children (calculated based on all relationship questions in the household). We define variables for unemployment, public assistance receipt and wages from detailed ACS questions on income and labor force participation. Finally, we define several human capital attainment variables – being below-grade-for-age, high school non-completion and college attendance – from detailed ACS questions on school attendance and highest grade completed.

The ACS also provides socio-demographic information for the second generation, including race, sex, and age. Since the first generation do not always appear in the ACS at the same time as the second generation, we attach demographic characteristics from the decennial Census to the second generation. We also collect information on the characteristics of the first generation's county of birth – population, employment, personal income per capita and total transfer income – from the Bureau of Economic Analysis' Regional Product Accounts.

3 Empirical Strategy

3.1 Baseline Econometric Model

We are interested in estimating the relationship between airborne particulate matter exposure in early childhood and the later life outcomes of affected individuals around 30 years later, alongside their children – the second generation. Our empirical specification makes use of the same structure and controls as Isen et al. (2017). Our baseline model takes the following form,

$$Outcome_{j,c,t} = \beta_0 + \beta_1 TSP_{j,c,t} + \gamma X_j' + \lambda X_c' t + \alpha_c + \alpha_{st} + \epsilon_{j,c,t,y} \tag{1}$$

where $TSP_{j,c,t}$ is the average particulate matter concentration that individual j was exposed to in county c and year t, measured in $\mu g/m^3$. X_j is a vector of individual characteristics, including age, race, and sex, as well as in-utero weather exposure. $X_c t$ is a vector of pre-1970 CAAA county-level characteristics interacted with linear and quadratic time trends. α_c are county-of-birth fixed effects that control for time-invariant unobserved determinants of the labor market outcomes and family structure for individuals born in county c. α_{st} are birth-state $\times$ birth-year fixed effects which control for time-varying determinants of the long-run outcomes, common across all individuals born in a state s in year t. The coefficient of interest is β_1 which estimates the effect of a one-unit increase in TSP emissions on an individual's later life outcomes, holding constant individual demographic characteristics, pre-CAAA 1970 county-of-birth characteristics, and county-of-birth weather patterns in utero.

In evaluating the effects of the CAAA on the second generation – we use a comparable empirical specification,

$$Outcome_{i,j,c,t} = \beta_0 + \beta_1 TSP_{j,c,t} + \gamma X_j' + \lambda X_c' t + \alpha_c^j + \alpha_{st}^j + \epsilon_{i,c,t,y} \tag{2}$$

Our main specification is an exact replication of equation 1, containing only parental variables (subscripted by j). However, in extensions we allow for additional second-generation controls, including birth-county fixed effects, α_c^i, birth-state $\times$ birth-year fixed effects, α_{st}^i, and second-generation individual characteristics, X_i' in addition to the first generation fixed effects and characteristics. Across all specifications we cluster our standard errors at the first generation county of birth level.

It is highly likely that exposure to particulate matter is correlated with many observable and unobservable determinants of long-run economic and social outcomes. While the inclusion of birth-county and birth-state $\times$ birth-year fixed effects will absorb any time-invariant county-specific determinants and time-varying determinants common to all individuals in a given state-year, it is likely that individual-level or local-level factors that correlate with particulate matter still exist, leading to bias in our OLS estimates of β_1.

3.2 Using the 1970 CAAA in an Instrumental Variables Design

To address the endogeneity concerns related to pollution exposure, we instrument for changes in particulate matter exposure using the introduction of the 1970 Clean Air Act Amendment. The Clean Air Act was introduced in 1963 and regulates air pollution in the United States and is the largest environmental program in the country. It requires the EPA to develop and enforce regulations to protect the population from exposure to airborne pollutants that are known to be hazardous to human health. In 1970 the Clean Air Act was amended, authorizing federal regulations to limit emissions, resulting in a major shift in the federal government's role in air pollution control. As a consequence of the 1970 amendments the EPA established the national ambient air quality standards (NAAQS), specifying the minimum level of air quality that is acceptable for six criteria air pollutants – Sulfur dioxide (SO_2), particulate matter (TSP, $PM_{2.5}$ and PM_{10}), nitrogen dioxide (NO_2), carbon monoxide (CO), Ozone (O_3), and lead. Areas that exceeded these standards were considered to be non-attainment areas. As a consequence of being designated a non-attainment area a plan must be implemented to meet the standard. If this is not put into place then the area risks losing federal financial assistance.

Prior research has shown that non-attainment designations resulted in substantial reductions in TSP concentrations (Henderson, 1996). Chay and Greenstone (2003a, 2005) then used these regulatory-induced changes in particulate concentrations to better understand the relationship between particulate matter exposure infant health, and the willingness to pay for air quality more generally, documenting significant reductions in infant mortality. We explore whether these same changes in air pollution have long-run effects on the labor market outcomes and family structure of the cohort's who survived, and the degree to which these changes affected the second generation – the children of the cohort's who survived.

We model the change in air pollution using an indicator variable for county non-attainment status interacted with an indicator for the years 1972 or later. The first stage regression in this two-stage least squares estimator is essentially a difference-in-difference regression model,

$$TSP_{i,c,t} = \alpha_0 + \alpha_1(Nonattainment_{c,1970} \times 1[\tau > 1971]) + \eta_c + \eta_{st} + \gamma X_i' + \delta X_c' t + \nu_{ict} \quad (3)$$

where TSP exposure for individual i in county c in year t is regressed on a time-invariant county indicator equal to 1 if a county is designated as non-attainment, $Nonattainment_{c,1970}$, interacted with an indicator equal to 1 for the years after the CAAA went into affect, $1[\tau > 1971]$. The interaction term is therefore equal to 1 for non-attainment counties following the implementation of the 1970 CAAA. The parameter of interest is α_1, which provides a difference-in-difference estimate of the impact of non-attainment designation on in-utero TSP exposure in the years after CAAA regulations went into place.

In the second stage, we use the predicted TSP levels from equation 3 in place of observed TSP levels,

$$Outcome_{i,c,t} = \rho_0 + \rho_1 \widehat{TSP}_{i,c,t} + \eta_c + \eta_{st} + \gamma X_i' + \delta X_c' t + \varepsilon_{j,c,t,y} \tag{4}$$

where the coefficient of interest ρ_1 captures the effect of a one-unit increase in CAAA-driven TSP within an individual's birth-county and birth-year on their later life outcomes holding constant individual demographic characteristics, pre-1970 CAAA county-of-birth characteristics, and county-of-birth, in-utero, weather exposure.

We show, consistent with previous research on the Clean Air Act, that the first stage relationship is strong – that non-attainment designation is associated with significant and persistent declines in particulate matter concentrations in the years after the 1970 CAAA came into effect. Figure 2 summarizes the effects of nonattainment on TSP in utero exposure in an event study framework, where nonattainment has separate effects in each year. The pre-1972 effects can be interpreted as placebo tests and provide a test for the parallel trends assumption. We find statistically, and economically, insignificant effects prior to 1972 – evidence in support of our identification strategy – and large declines in TSP exposure following 1972 – evidence of instrument relevance.

In addition, Isen et al. (2017) provide evidence to suggest that the instrument may satisfy the exclusion restriction required for consistent estimation of ρ_1. Specifically, there are limited differences between attainment and non-attainment county characteristics prior to the 1970 CAAA, providing further support for the parallel trends assumption. In addition, non-attainment designation does not appear to be correlated with changes to the observable characteristics of mothers that gave birth in the years following the 1970 CAAA.

That being said we can never rule out the possibility that the exclusion restriction may be violated as the CAAA may have affected counties in ways other than through reductions in pollution. Isen et al. (2017) make the point that non-attainment designations affected economic competitiveness (Greenstone, 2002; Greenstone et al., 2012; Walker, 2011, 2013). However, the effects on the broader local economy are small affecting less than 0.7 percent of the total workforce (Walker, 2013) and based on the 1990 Clean Air Act Amendments. As the

1970 CAAA was the first major regulation to take place actions to reduce emissions may have been less costly than in the 1990s, attenuating concerns about economic competitiveness. Nevertheless, we cannot rule out the potential that the 1970 CAAA contributed to a decline in economic conditions for non-attainment counties, affecting the long-run economic prospects of affected individuals. In Appendix A we report the reduced form effects of non-attainment, which capture the overall effects of the CAAA on individuals born into non-attainment counties, following the introduction of the 1970 CAAA.

4 Results

We summarize our results in two stages. First we analyze the effect of pollution exposure at birth on the later-life outcomes of individuals that were directly affected by the 1970 CAAA – the first-generation effect. Second, we examine whether the direct effects of the 1970 CAAA are transmitted across generations to affect the human capital of the second generation – the children of those that were in-utero exposed.

4.1 First Generation Outcomes

We begin by exploring how the large reductions in pollution exposure induced by 1970 CAAA affected outcomes for individuals born between 1960-1980. We evaluate the long-run impacts using their responses to the ACS as adults. We consider two main sets of outcomes: family structure outcomes, and labor market outcomes. Except where noted, the estimating sample for each set of regressions includes only individuals identified as parents described in the data section above.

We begin by examining the effects of in-utero TSP exposure on later life labor market outcomes. Table 2 presents the findings of this exercise. Our outcomes of interest are whether the affected individual is unemployed at the time of the ACS survey response, whether they are, or have been, in receipt of public assistance during the previous 12 months, and what their current earnings are, conditional on being employed. Across both the OLS and IV estimates we find no effects of in-utero TSP exposure on the likelihood of being unemployed or in receipt of public assistance later in life, suggesting that this exposure does not appear to affect labor market outcomes on the extensive margin per se.[11] However, we do find significant effects of in-utero TSP exposure on wages in the IV specifications. A 10 $\mu g/m^3$ reduction in in-utero exposure to TSP is associated with a 0.9 percent increase in earnings at the time of ACS response. Evaluated at the mean earnings, this corresponds to an annual earnings effect of $407. This effect is similar in magnitude to the estimates presented in Isen et al. (2017), which is encouraging given the we are using the same methodology. Furthermore, it is encouraging to note the external validity of the Isen et al. (2017) findings, that are restricted to 26 states. Our sample contains individuals in all states which had active TSP monitors in 1970, and does not restrict the age of respondents to 29-31.

Assuming that this effect is constant over the life cycle, and that earnings are discounted at a real rate of 3 percent back to age zero (5 percent discount rate + 2 percent wage growth), the lifetime earnings effect of a 10 $\mu g/m^3$ increase in TSP is $4,066.11 per person. Furthermore, if we assume a linear dose response function and apply the observed changes in TSP since 1971, the aggregate increase in earnings associated with reduced in-utero TSP exposure come to $1.978 trillion.

In addition to examining the effects on economic outcomes we also explore the effects of in-utero exposure to TSP on family structure. Our outcomes of interest are whether an affected individual is divorced at the time of the survey, whether they are married, whether they have any children, the number of children that they have, and the whether the first child was conceived as a teenager. We are interested in exploring these outcomes for two reasons. First, because we believe that family structure could plausibly affect parental investments and resources available to children, which could affect the intergenerational transmission of human capital. Secondly, we are interested in the potential for selection in the second

[11] This is consistent with Isen et al. (2017) who estimate increases in the number of quarters worked, rather than labor force participation effects, for affected individuals in 24 states in the LEHD dataset.

generation sample, if pollution affected fertility. Table 3 presents the results of these findings.

First, we find that there are small changes in family structure through the likelihood of divorce in the IV specification. A 10 $\mu g/m^3$ increase in in-utero exposure to TSP is associated with a 0.2 percent increase in the likelihood of being divorced. However, this effect is very small and there appear to be no corresponding changes in the likelihood of being married.

Second, we find that there are no statistically significant effects on the likelihood of having any children, the number of children, or the timing of children. This finding is important as it reduces the likelihood that there is selection based on fertility into the second generation sample. Furthermore, it contributes to the literature on income and fertility, in which research has explored whether children are normal or inferior goods (Lindo, 2010; Black et al., 2013). Our findings suggest that demand for children is relatively inelastic; however, it is important to caveat that the lifetime income effects are not huge and so may not be sufficient to drive a fertility response on the margin.

In Appendix A we present results showing the robustness of our findings to the application of a regression discontinuity design approach as an alternative first stage – exploiting variation for counties around the non-attainment threshold – as well as the imposition of various sample restrictions.

4.2 Second Generation Outcomes

In light of our findings above, it is of interest to explore the intergenerational consequences of the 1970 CAAA. We examine how the large reduction in pollution exposure experienced in-utero by the first-generation born in non-attainment counties, following the 1970 Clean Air Act Amendments affected the human capital acquisition of their children – the second generation. Our focus on human capital is largely necessitated by timing – very few individuals over the age of 22 at ACS response have parents born after 1971, limiting the sample available to examine the effects on wages and labor force participation difficult. However, human capital measures available in the ACS which are prevalent among people under 22, who are much more likely to have "treated" parents.

Examining the effects of a parent's pollution exposure on their child yields an additional layer of complexity to our analysis. The regression tables summarize four specifications we will use to understand this complexity, starting with an identical specification to the parent effects regressions. Following this main specification we then incorporate child characteristics, and finally include children's county of birth and state-by-year of birth fixed effects. These additional controls are helpful to address any avoidance-related intergenerational sorting, which in turn sheds light on the degree to which economic behavior can affect the intergenerational transmission of pollution exposure.

Table 5 presents results examining the second-generation effects of in-utero TSP exposure on the likelihood that the second-generation child is below their grade-for-age. Across all specifications we find limited evidence that this is the case. Similarly, Table 6 explores the likelihood that a second-generation child drops out of high school, again resulting in null effects across all OLS and IV specifications. Collectively, these findings suggest that the intergenerational transmission of pollution exposure may not have an important effect on education in the earlier stages of the life cycle.

However, in Table 7 we find that there are significant effects of in-utero pollution exposure on the likelihood that the second generation attends college.[12]

We estimate that a 10 $\mu g/m^3$ reduction in first generation in-utero pollution exposure is associated with a 2 percentage point increase in the likelihood that a second generation child attends college. These estimated effects on college attendance likely have substantial impacts on the earnings potential of second-generation individuals. Using a college wage premium of $25,000 p.a. (estimated using Mincerian wage regression), combined with a graduation rate of 50 percent to convert college attendance to college completion (NCES), and a 3 percent real discount rate we predict a second generation lifetime earnings at age

[12]It is important to note that this is a selected sample, relative to the second generation as whole as these are likely first-born children born to young parents. This may affect the degree to which our findings generalize to the rest of the population. However, we do not observe any observable differences between this sample and younger second generation children (other than age).

zero (of the first generation) equal to $\$703/10\mu g/m^3$.[13] Combining the number of births since 1971, the probability that an individual has a child (0.61), and the average number of children had by each individual (1.4), the aggregate benefits of observed TSP reductions since 1971 to the second generation amount to $292 billion.

Combining the first and second generation estimates the aggregate earnings impact of in-utero pollution exposure is 15 percent higher than our estimates based on a within-generation analysis, suggesting that a large part of the pollution exposure effect on earnings is likely to be passed on to the second generation. These estimated effects on college attendance are substantial. To provide some context for the magnitude of the effect, we draw on the existing literature, which has explored the determinants of college attendance.

4.3 Understanding Mechanisms

4.3.1 Income and Parental Resources

The absence of an effect of grade-for-age and high school completion, but large effects on college suggest a mechanism that may be of specific relevance to college attendance. One consideration is that increases in parental income as shown above, reduce liquidity constraints; however, the estimated income effects imply a relationship which is much larger than the existing literature. Lovenheim (2011) and Lovenheim and Lockwood Reynolds (2013) explore the effects of an increase in household wealth on college attendance finding that a $10,000 increase in housing wealth is associated with a 0.71 - 0.92 percentage point increase in the likelihood of attending college. Bulman et al. (2017) explore the effects of winning the lottery on college attendance. They find that a $10,000 increase in housing wealth is associated with a 0.2 percentage point increase in attending college. By contrast, if our findings are driven entirely by the increase in parental earnings then a $10,000 increase in household wealth would be associated with a 4.91 percentage point increase in the

[13] If we discount back to age zero for the second generation we get a lifetime discounted earnings at age zero estimate of $2,379, 58 percent of the first generation effect.

likelihood of attending college. As such, we believe that the magnitude of the effect is too substantial to be driven entirely by the increase in parental income. One caveat with this interpretation is that the existing literature explores the effects of an increase in individual household wealth. By contrast, our effect captures an increase in wealth for the whole community. Consequently, there may be general equilibrium effects associated with this increase in wealth that contribute to college attendance decisions.

For example, Chay and Greenstone (2005) show that the 1970 Clean Air Act resulted in increases in house prices. This may affect the composition of the local economy through sorting, and resources available for schools as school funding is determined by local property taxes. However, we already control for county-of-birth fixed effects for the parent and in certain specifications control for county-of-birth fixed effects for children as well. Consequently, if children have not moved then county-level considerations will be captured already, i.e., our estimates are net of these considerations. As such we argue that the effect is likely to be driven by parental resources and investments rather than community-level considerations.[14]

4.3.2 Cognitive and Non-Cognitive Skills

An alternative explanation to a direct income effect is the consequences of increased health and resources for investments in the cognitive and non-cognitive skills of children (Murnane et al., 2000; Heckman and Carneiro, 2003; Belfield et al., 2006; Cunha et al., 2010; Heckman et al., 2013; Lundberg, 2017; Akee et al., 2018). If reduced exposure to pollution increases parental health as well as wealth, then parents may be better placed to spend time and make investments in their children – reductions in pollution may improve parental human capital. With the data available it is difficult to directly evaluate this considerations; however, again the implied magnitude of the effect based on the existing literature suggest that this channel

[14]One caveat to this is if the children of affected parents are more likely to have moved away from their county-of-birth, i.e., if county-of-residence is not the same as county-of-birth at the time of ACS response. We do not find any differential effect in the likelihood that non-attainment parents are more likely to have migrated that parents born into attainment counties; however, they may move to different places. Consequently, future work will explore this consideration.

is unlikely to fully explain the results. Belfield et al. (2006) explore the effects of the Perry Preschool program, which has been shown to have significant effects on childrens' cognitive, and especially non-cognitive skills. We estimate that a $10\mu g/m^3$ reduction in TSP has an equivalent effect on college attendance to 0.13 Perry Preschool Programs. Lundberg (2017) explores how specific non-cognitive and cognitive skills are associated with college attendance, based on her results, we find that a $10\mu g/m^3$ reduction in TSP is equivalent to a 1.14 standard deviation increase in self esteem, a 0.53 standard deviation reduction in impulsivity, a 0.4 standard deviation decrease in schooling problems, and a 0.2 standard deviation increase in cognitive ability.

An existing literature has explored the effects of pollution on crime, and school behavior, arguing that pollution increase impulsivity. Consequently, it is not outside the realms of possibility that improvements in pollution could improve school behavior and reduce criminal activity, contributing to increases in the likelihood of college attendance. However, again one might conceive that such a channel would also affect the likelihood of being a high school dropout.

One indirect approach to understanding the relevance that parental investments in cognitive and non-cognitive skills may play is to examine the relationship between pollution exposure and parental time-use. We do this by linking the American Time-Use Survey with our existing data infrastructure. This allows us to explore the effects of in-utero exposure for our first generation, on the time spent on activities with their children. We look at the effects of parental exposure on the time spent reading with children, and the time spent on educational activities. We find that reductions in gestational particulate matter exposure are associated with an increase in time spent reading to children. The coefficient is small in magnitude corresponding to an average increase of 0.14 minutes per day; however this is off a very low baseline mean of 1.4 minutes per day and so represents a roughly 10% increase in time spent reading.

Evidence suggests that reading to your children can help them to develop empathy, deal with difficult issues, improve vocabulary and background knowledge, increase attention span, and improve family relationships (Anderson et al., 1985; Koralek, 2003; Massaro, 2017; Mendelsohn et al., 2018). Consequently, the estimated college attendance effects may arise in part because of parental investments in cognitive and non-cognitive skills.

4.3.3 Genetic Transmission

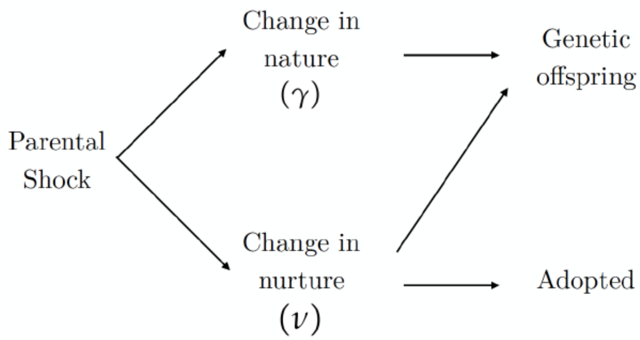
One margin that we are currently able to explicitly explore is the degree to which the effect is (broadly) driven by parental resources and investments as opposed to genetics. To do this we examine the differential effect of parental TSP exposure on adopted versus biological children. This exercise allows us to examine the degree to which genetic factors, passed down from parent to child may have affected ability and lowered the returns to college attendance. As the Decennial Census and ACS both ask whether the child of the head of household is natural born or adopted, we are able to identify a set of parent-child links for which there should be no direct genetic channel of transmission. By examining whether the effect of parental exposure systematically varies between adopted and non-adopted children, we can interrogate whether such a genetic channel may be important.

One may be concerned that there is a differential propensity to adopt or not adopt children in non-attainment counties, and so we first explore whether there is selection into the second-generation adopted sample. We do this by estimating our baseline first generation regressions, using an indicator for whether a parent has an adopted child as the dependent variable. Table 8 summarizes this result, finding no statistically effect on adoption for either TSP exposure (in OLS or IV specifications) or for nonattainment designations in the reduced form.

We then explore whether there is a differential effect of pollution on college attendance for adopted versus biological children. Figure 1 highlights the pathways through which a shock to parental endowments could flow differently to biological offspring and adopted offspring.

Biological offspring could be affected through both changes in household environment

Figure 1: The differential effect of a shock to parental endowments on biological and adopted children



(ν) and through genetic channels (γ). By contrast, adopted offspring can only be affected through the household environment (ν), i.e., parental resources and investments.[15]

If the effects on college attendance are entirely driven by genetic pathways ($\nu = 0$) then we would expect there to be no effect on adopted children, i.e, a differential effect that is the negative of the effect on biological children, $-\gamma$. If the college attendance effect is entirely driven by the household environment ($\gamma = 0$) then there should be no differential effect on adopted children, and the coefficient on the interaction term should be zero.

Of course, it is entirely possible that the college attendance effect is a mix of both channels $(1-\phi)\nu - (\phi\gamma + (1-\phi)\nu) = -\phi\gamma$ at which point the effect on adopted children should be smaller than the effect on biological children, unless parents differentially invest in adopted children, at which point the effect could be larger than the effect on biological children. We evaluate these considerations by estimating our IV specification, incorporating the interaction between parental TSP exposure and whether the child is adopted.

Across all specification we find no statistically different effects of parental pollution exposure between adopted and biological children, suggesting that the estimated college attendance effect arises due to parental resources and investments, rather than genetic pathways

[15]One caveat to this analysis is that adopted children may be affected genetically if their biological parents were born in non-attainment counties at the same time as their adopted parents. However, this requires that their adopted parents and biological parents are born at the same time and location, which we posit is unlikely to be the case in a systematic way.

(Bjorklund and Chadwick, 2003; Bjorklund, 2006; Bjorklund et al., 2007, 2010). However, the magnitude of the effects provide suggestive evidence that there may even be differential investments towards adopted children. Taken at face value the estimated effect on adopted children is 2.9-3.7 percentage points, compared to a 2.1 - 2.3 percentage point increase for biological offspring. This is not to say that within-household parents invest more in their adopted children. Instead, on average adopted children experience an increased likelihood of attending college if their adopted parents were born exposed to reductions in pollution than children who's adopted parents were exposed to higher levels of pollution.[16]

5 Conclusion

In this paper we provide the first quasi-experimental evidence of the intergenerational consequences of in-utero exposure to ambient air pollution. Exploiting variation in particulate matter, induced by the introduction of the 1970 Clean Air Act amendments, which substantially reduced ambient air pollution, we find that individuals that were directly exposed experienced increased earnings and more stable family environments, through a reduction in the likelihood of divorce 25-35 years later.

In addition, we estimate that this reduction in pollution exposure has an effect on their children – the second generation – through an increase in the likelihood that they attend college, highlighting the importance of environmental factors for the intergenerational transmission of human capital. Collectively, we estimate that a combined first- and second-generation estimate on aggregate earnings is 17-58 percent higher than estimates based solely on the first generation. As such, within-generation estimates of the dose-response function likely underestimate the total welfare effects of environmental toxins.

We also provide some insights into the underlying mechanism through which this shock to parental endowments affects economic opportunity for the second generation. Using infor-

[16]Unfortunately, we do not have sufficient power to evaluate the effects within-household as it would require siblings to both be in the ACS at the same time.

mation on the adopted status of children, we explore whether differential effects of parental pollution exposure, seeking to highlight the importance of genetic factors, as opposed to parental resources and investments. We estimate that there is no differential effect between biological and adopted children, nor any effect on the propensity of the first generation to adopt. This points to the importance of parental resources and investments as the underlying mechanism for the transmission of human capital across generations in this context.

Our findings are additionally relevant for the literature on environmental justice. The focus of this debate has been on equity considerations – whether disadvantaged populations also bear disproportionate burden of environmental harms. However, our results highlight that there are also efficiency consequences of environmental inequality. The interests of future generations are not capitalized into market outcomes, despite evidence to suggest that they are affected by the exposure of their parents today. Consequently, improvements in air quality could play a role in reducing economic inequality both across space and generations.

It is striking that these effects arise from such short exposure times, early in life, and that the effects appear to be driven by parental investments and resources, rather than genetics. As such, our results suggest that Pareto improvements could be made by re-allocating resources from later to earlier in the life-cycle. Understanding, the margins through, and degree to, which such reallocation can be delivered is an important area for future research.

References

Akee, R., W. Copeland, E. Costello, and E. Simeonova (2018): "How Does Household Income Affect Child Personality Traits and Behaviors?" *American Economic Review*, 108, 775–827.

Akresh, R., D. Halim, and M. Kleemans (2018): "Long-Term and Intergenerational Effects of Education: Evidence from School Construction in Indonesia," *NBER Working Paper 25265*.

Almond, D. (2006): "Is the 1918 Influenza Pandemic Over? Long-Term Effects of In-Utero Influenza Exposure in the Post-1940 U.S. Population," *Journal of Political Economy*.

Almond, D., L. Edlund, H. Li, and J. Zhang (2010): "Long-Term Effects of the 1959-61 China Famine: Mainland China and Hong Kong," in *The Economic Consequences of Demographic Change in East Asia*, ed. by T. Ito and A. Rose, NBER-EASE Vol. 19.

Anderson, R., E. Hiebert, J. Scott, and L. Wilkinson (1985): "Becoming a Nation of Readers: The Reprot of the Commission on Reading," Tech. rep., U.S. Department of Education.

Barr, A. and C. Gibbs (2017): "Breaking the Cycle? The Intergenerational Effects of Head Start," *Mimeo*.

Becker, G. and N. Tomes (1979): "An Equilibrium Theory of the Distribution of Income and Intergenerational Mobility," *Journal of Political Economy*.

Belfield, C., M. Nores, S. Barnett, and L. Schwienhart (2006): "The High/Scope Perry Preschool Program: Cost-Benefit Analysis Using Data from the Age-40 Followup," *Journal of Human Resources*.

Bjorklund, A. (2006): "The Origins of Intergenerational Associations: Lessons from Swedish Adoption Data," *Quarterly Journal of Economics*, 121, 999–1028.

Bjorklund, A. and L. Chadwick (2003): "Intergenerational Income Mobility in Permanent and Separated Families," *Economic Letters*, 80, 239–246.

Bjorklund, A., M. Jantti, and G. Solon (2007): "Nature and Nurture in the Intergenerational Transmission of Socioeconomic Status: Evidence from Swedish Children and Their Biological and Rearing Parents," *B.E. Journal of Economic Analysis and Policy: Advances in Economic Analysis and Policy*, 7.

Bjorklund, A., L. Lindahl, and M. Lindquist (2010): "What More Than Parental Income? An Exploration of What Swedish Siblings Get from Their Parents," *B.E. Journal of Economic Analysis and Policy: Contributions to Economic Analysis and Policy*, 10.

Black, D., N. Kolesnikova, S. Sanders, and L. Taylor (2013): "Are Children "Normal"?" *Review of Economics and Statistics*, 95, 21–33.

Black, S., A. Butikofer, P. Devereux, and K. Salvanes (2018): "This is Only a Test? Long-Run Impacts of Prenatal Exposure to Radioactive Fallout," *Review of Economics and Statistics*.

Black, S. and P. Devereux (2011): "Recent Developments in Intergenerational Mobility," in *Handbook of Labor Economics*, Elsevier.

Black, S., P. Devereux, and K. Salvanes (2005): "Why the Apple doesn't fall far: Understanding Intergenerational Transmission of Human Capital," *American Economic Review*.

——— (2007): "From the Cradle to the Labor Market? The Effect of Birth Weight on Adult Outcomes," *Quarterly Journal of Economics*.

Bulman, G., R. Fairlie, S. Goodman, and A. Isen (2017): "Parental Resources and College Attendance: Evidence from Lottery Wins," *Mimeo*.

Chay, K., C. Dobkin, and M. Greenstone (2003): "The Clean Air Act of 1970 and Adult Mortality," *Journal of Risk and Uncertainty*.

Chay, K. and M. Greenstone (2005): "Does Air Quality matter? Evidence from the Housing Market," *Journal of Political Economy*.

Chay, K. Y. and M. Greenstone (2003a): "Air Quality, Infant Mortality, and the Clean Air Act of 1970," Working Paper 10053, National Bureau of Economic Research.

——— (2003b): "The Impact of Air Pollution on Infant Mortality: Evidence from Geographic Variation in Pollution Shocks Induced by a Recession," *The Quarterly Journal of Economics*, 118, 1121–1167.

Chetty, R., J. Friedman, N. Hendren, M. Jones, and S. Porter (2018): "The Opportunity Atlas: Mapping the Childhood Roots of Social Mobility," *NBER Working Paper 25147*.

Chetty, R. and N. Hendren (2018a): "THe Effects of Neighborhoods on Intergenerational Mobility I: Childhood Exposure Effects," *Quarterly Journal of Economics*.

——— (2018b): "THe Effects of Neighborhoods on Intergenerational Mobility II: County Level Estimates," *Quarterly Journal of Economics*.

Chetty, R., N. Hendren, P. Kline, and E. Saez (2014): "Where is the land of Opportunity? The Geography of Intergenerational Mobility in the United States *," *The Quarterly Journal of Economics*, 129, 1553–1623.

Cunha, F., J. J. Heckman, and S. M. Schennach (2010): "Estimating the Technology of Cognitive and Noncognitive Skill Formation," *Econometrica*, 78, 883–931.

Currie, J. (2011): "Inequality at Birth: Some Causes and Consequences," *Ely Lecture - American Economic Review*.

Currie, J., M. Neidell, and J. Schmieder (2009): "Air Pollution and Infant Health: Lessons from New Jersey," *Quarterly Journal of Economics*.

East, C., S. Miller, M. Page, and L. Wherry (2017): "Multi-Generational Impacts of Childhood Access to the Safety Net: Early Life Exposure to Medicaid and the Next Generation's Health," *NBER Working Paper 23810*.

Fertig, A. and T. Watson (2009): "Minimum Drinking Age Laws and Infant Health Outcomes," *Journal of Health Economics*.

Greenstone, M. (2002): "The Impacts of Environmental Regulations on Industrial Activity: Evidence from the 1970 and 1977 Clean Air Act Amendments and the Census of Manufactures," *Journal of Political Economy*.

Greenstone, M., J. List, and C. Syverson (2012): "The Effects of Environmental Regulation on the Competitiveness of US Manufacturing," *Mimeo*.

Gruber, J. (2004): "Is Making Divorce Easier Bad for Children? The Long-Run Implications of Unilateral Divorce," *Journal of Labor Economics*.

Heckman, J. and P. Carneiro (2003): "Human Capital Policy," *NBER Working Paper 9495*.

Heckman, J., R. Pinto, and P. Savelyev (2013): "Understanding the Mechanisms Through Which an Influential Early Childhood Program Boosted Adult Outcomes," *American Economic Review*, 103, 2052–2086.

Henderson, J. (1996): "Effects of Air Quality Regulation," *American Economic Review*.

Hsiang, S., P. Oliva, and R. Walker (2018): "The Distribution of Environmental Damages," *Review of Environmental Economics and Policy*.

Isen, A., M. Rossin-Slater, and W. R. Walker (2017): "Every Breath You Take — Every Dollar You'll Make: The Long-Term Consequences of the Clean Air Act of 1970," *Journal of Political Economy*.

Kelly, E. (2011): "The Scourge of Asian Flu: In Utero Exposure to Pandemic Influenza and the Development of a Cohort of British Children," *Journal of Human Resources*.

Koralek, D. (2003): "Reading Aloud with Children of All Ages," Tech. rep., Reading Is Fundamental.

Lindo, J. (2010): "Are Children Really Inferior Goods? Evidence from Displacement-Driven Income Shocks," *Journal of Human Resources*, 45, 301–327.

Lovenheim, M. (2011): "The Effect of Liquid Housing Wealth on College Enrollment," *Journal of Labor Economics*, 29, 741–771.

Lovenheim, M. and C. Lockwood Reynolds (2013): "The Effect of Housing Wealth on College Choice: Evidence from the Housing Boom," *Journal of Human Resources*, 14, 3–37.

Lundberg, S. (2017): "Non-Cognitive Skills as Human Capital," in *Education, Skills, and Technical Change: Implications for Future U.S. GDP Growth*, ed. by C. Hulten and V. Ramey, Cambridge: National Bureau of Economic Research, 219–243.

Massaro, D. W. (2017): "Reading Aloud to Children: Benefits and Implications for Acquiring Literacy Before Schooling Begins," *The American Journal of Psychology*, 130, 63–72.

Mendelsohn, A., C. Brockmeyer, A. Weisleder, S. Berkule Johnson, A. Seery, C. Canfield, H. Huberman, and B. Dreyer (2018): "Reading Aloud, Play, and Social-Emotional Development," *Pediatrics*, 141.

Murnane, R. J., J. B. Willett, Y. Duhaldeborde, and J. H. Tyler (2000): "How important are the cognitive skills of teenagers in predicting subsequent earnings?" *Journal of Policy Analysis and Management*, 19, 547–568.

Sanders, N. and C. Stoecker (2015): "Where have all the young men gone? Using Sex Ratios to Measure Fetal Death Rates," *Journal of Health Economics*.

Solon, G. (1992): "Intergenerational Income Mobility in the United States," *American Economic Review*.

Voorheis, J. (2017a): "Air Quality, Human Capital Formation and the Long-term Effects of Environmental Inequality at Birth," Working Paper 2017-05, US Census Bureau.

——— (2017b): "Longitudinal Environmental Inequality and Environmental Gentrification: Who Gains From Cleaner Air?" Working Paper 2017-04, US Census Bureau.

Wagner, D. and M. Layne (2014): "The Person Identification Validation System (PVS): Applying the Center for Administrative Records Research and Applications (CARRA) Record Linkage Software," *Mimeo*.

Walker, R. (2011): "Environmental Regulation and Labor Reallocation," *American Economic Review: Papers and Proceedings*.

——— (2013): "The Transitional Costs of Sectoral Reallocation: Evidence from the Clean Air Act and the Workforce," *Quarterly Journal of Economics*.

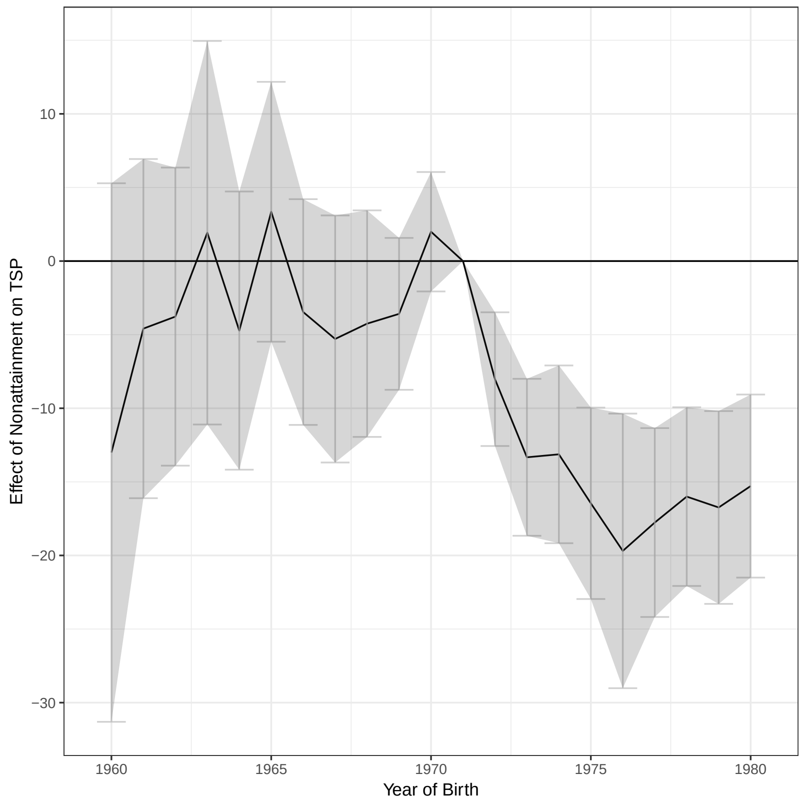
Tables and Figures

Table 1: First Generation Descriptive Statistics by Treatment

	(1) Attainment County	(2) Non-Attainment County
Female	0.550	0.555
	(0.498)	(0.497)
Black	0.065	0.095
	(0.247)	(0.293)
White	0.778	0.735
	(0.416)	(0.442)
Hispanic	0.061	0.086
	(0.239)	(0.280)
Other	0.095	0.084
	(0.294)	(0.277)
Age	41.76	41.99
	(6.506)	(6.473)
Gestational Exposure (10 $\mu g/m^3$)	7.301	10.77
	(2.761)	(3.775)
County Population	405,000	1,399,000
	(445,000)	(1,923,000)
Personal Income per Capita	3,391	4,229
	(780.3)	(847.1)

Notes: Source: Census Numident, Decennial Census Short Form 2000 and 2010, ACS 2005 through 2015. Approved for release by the Census DRB, authorization numbers CBDRB-FY18-232, CBDRB-FY18-248, CBDRB-FY18-287 and CBDRB-FY19-099.

Figure 2: First Stage: The Effect of Nonattainment Designations on TSP Exposure In Utero



Source: Census Numident, Decennial Census Short Form 2000 and 2010, ACS 2005 through 2015. Each point represents the coefficient on the interaction between an individual's county-of-birth nonattainment status and the year in question from a regression which also contains demographic controls including sex, race/ethnicity and quadratic in age, pre-CAA 1970 county of birth economic characteristics (employment, total transfer income, personal income per capita) interacted with quadratic trends, county of birth weather controls including average and maximum temperature and number of precipitation days during an individual's 9 month gestational period. Approved for release by the Census DRB, authorization numbers CBDRB-FY18-232, CBDRB-FY18-248, CBDRB-FY18-287 and CBDRB-FY19-099.

Table 2: The Long-run Effects of In-Utero TSP Exposure on Labor Market Outcomes

	(1) Unemployed	(2) Public Assistance	(3) log(Wages)	(4) log(Wages) (Below Median)
Panel A: OLS estimates				
TSP exposure ($10\mu g/m^3$)	0.00006	0.00002	−0.00050	−0.00047
	(0.00007)	(0.00004)	(0.00034)	(0.00048)
Panel B: IV estimates				
TSP exposure ($10\mu g/m^3$)	0.00032	−0.00085	−0.00873**	−0.01134**
	(0.00072)	(0.00061)	(0.00423)	(0.00472)
Fixed Effects	County-of-birth, State-of-birth × Year, Birth Month			
Individual Controls	Yes	Yes	Yes	Yes
County-level Controls	Yes	Yes	Yes	Yes
Control Mean	0.048	0.020	\$45,320	\$25,870
Observations	4,766,000	4,767,000	3,391,000	1,475,000
First Stage F-stat	21.52	21.53	21.87	23.47

Notes: Significance levels are indicated as * 0.10 ** 0.05 *** 0.01. Source: Census Numident, Decennial Census Short Form 2000 and 2010, ACS 2005 through 2015. All regressions include individual demographic controls including sex, race/ethnicity and quadratic in age, pre-CAA 1970 county of birth economic characteristics (employment, total transfer income, personal income per capita) interacted with quadratic trends, county of birth weather controls including average and maximum temperature and number of precipitation days during an individual's 9 month gestational period. Standard errors are clustered at the parent's county of birth level. Approved for release by the Census DRB, authorization numbers CBDRB-FY18-232, CBDRB-FY18-248, CBDRB-FY18-287 and CBDRB-FY19-099.

Table 3: The Long-run Effects of In-Utero TSP Exposure on Family Structure Outcomes

	(1) Divorced	(2) Married	(3) Any Kids	(4) # Kids	(5) Teen
Panel A: OLS estimates					
TSP exposure ($10\mu g/m^3$)	-0.00010	0.00011	0.00009	0.00039	0.0001
	(0.00010)	(0.00013)	(0.00025)	(0.00058)	(0.0001)
Panel B: IV estimates					
TSP exposure ($10\mu g/m^3$)	0.00225*	0.00008	0.00347	0.01680	-0.0015
	(0.00119)	(0.00163)	(0.00435)	(0.01127)	(0.0012)
Fixed Effects	County-of-birth, State-of-birth × Year, Birth Month				
Individual Controls	Yes	Yes	Yes	Yes	Yes
County-level Controls	Yes	Yes	Yes	Yes	Yes
Control Mean	0.125	0.72	0.61	1.410	
Observations	5,289,000	5,289,000	3,040,000	5,855,000	4,773,000
First Stage F-stat	21.55	21.55	19.09	19.23	

Notes: Significance levels are indicated as * 0.10 ** 0.05 *** 0.01. Source: Decennial Census Short Form 2000 and 2010, ACS 2005 through 2015. All regressions include individual demographic controls including sex, race/ethnicity and quadratic in age, pre-CAA 1970 county of birth economic characteristics (employment, total transfer income, personal income per capita) interacted with quadratic trends, county of birth weather controls including average and maximum temperature and number of precipitation days during an individual's 9 month gestational period. Standard errors are clustered at the parent's county of birth level. Approved for release by the Census DRB, authorization numbers CBDRB-FY18-232, CBDRB-FY18-248, CBDRB-FY18-287 and CBDRB-FY19-099.

Table 4: Second Generation Descriptive Statistics by Treatment

	(1) None Affected	(2) Mom Affected	(3) Dad Affected	(4) Both Affected
Female	0.487	0.489	0.486	0.486
	(0.500)	(0.500)	(0.500)	(0.500)
Black	0.089	0.146	0.073	0.093
	(0.285)	(0.353)	(0.261)	(0.291)
White	0.800	0.717	0.797	0.798
	(0.400)	(0.451)	(0.402)	(0.402)
Hispanic	0.092	0.121	0.113	0.097
	(0.289)	(0.327)	(0.317)	(0.297)
Other	0.018	0.016	0.016	0.012
	(0.134)	(0.125)	(0.126)	(0.107)
Age	12.16	13.33	11.49	11.74
	(7.302)	(7.605)	(7.413)	(6.89)
Gestational Exposure (10 $\mu g/m^3$)	5.162	5.532	5.255	5.484
	(1.667)	(1.796)	(1.733)	(1.712)

Notes: Source: Census Numident, Decennial Census Short Form 2000 and 2010, ACS 2005 through 2015. Approved for release by the Census DRB, authorization numbers CBDRB-FY18-232, CBDRB-FY18-248, CBDRB-FY18-287 and CBDRB-FY19-099.

Table 5: The Effect of In-Utero TSP Exposure on the Likelihood that the Second Generation are Below Grade-for-Age

	(1)	(2)	(3)	(4)	(5)
	OLS	IV			
	< Grade for Age	< Grade for Age	< Grade for Age	< Grade for Age	< Grade for Age
TSP exposure ($10\mu g/m^3$)	0.00004	-0.00101	-0.00086	-0.00091	-0.00035
	(0.00008)	(0.00093)	(0.00084)	(0.00083)	(0.00077)
Dep. Var. Mean	0.28	0.28	0.28	0.28	0.28
Observations	2,626,000	2,626,000	2,626,000	2,582,000	2,582,000
First Stage F-stat	–	26.44	26.44	26.29	26.31
1st Gen Fixed Effects	County-of-birth, State-of-birth × Year, Birth Month				
1st Gen Controls	Yes	Yes	Yes	Yes	Yes
2nd Gen Controls	No	No	Yes	Yes	Yes
2nd Gen County-of-birth FE?	No	No	No	Yes	Yes
2nd Gen State-of-birth×Year FE?	No	No	No	No	Yes

Notes: Significance levels are indicated as * 0.10 ** 0.05 *** 0.01. Source: Decennial Census Short Form 2000 and 2010, ACS 2005 through 2015. All regressions include the same set of control variables as in Table 22. Column 3-5 contain additional second generation demographic characteristics including race/ethnicity, sex and a quadratic in age. Standard errors are clustered at the parent's county of birth level. Approved for release by the Census DRB, authorization numbers CBDRB-FY18-232, CBDRB-FY18-248, CBDRB-FY18-287 and CBDRB-FY19-099.

Table 6: The Effect of In-Utero TSP Exposure on the Likelihood that the Second Generation do not Complete High School

	(1)	(2)	(3)	(4)	(5)
	OLS	IV			
	High School Dropout	High School Dropout	High School Dropout	High School Dropout	High School Dropout
TSP exposure ($10\mu g/m^3$)	0.00017	−0.00040	−0.00122	−0.00138	−0.00123
	(0.00011)	(0.00204)	(0.00209)	(0.00210)	(0.00206)
Dep. Var. Mean	0.06	0.06	0.06	0.06	0.06
Observations	1,252,000	1,252,000	1,252,000	1,252,000	1,252,000
First Stage F-stat	–	23.82	23.83	23.7	23.45
1st Gen Fixed Effects	County-of-birth, State-of-birth × Year, Birth Month				
1st Gen Controls	Yes	Yes	Yes	Yes	Yes
2nd Gen Controls	No	No	Yes	Yes	Yes
2nd Gen County-of-birth FE?	No	No	No	Yes	Yes
2nd Gen State-of-birth×Year FE?	No	No	No	No	Yes

Notes: Significance levels are indicated as * 0.10 ** 0.05 *** 0.01. Source: Decennial Census Short Form 2000 and 2010, ACS 2005 through 2015. All regressions include the same set of control variables as in Table 22. Column 3-5 contain additional second generation demographic characteristics including race/ethnicity, sex and a quadratic in age. Standard errors are clustered at the parent's county of birth level. Approved for release by the Census DRB, authorization numbers CBDRB-FY18-232, CBDRB-FY18-248, CBDRB-FY18-287 and CBDRB-FY19-099.

Table 7: The Effect of In-Utero TSP Exposure on the Likelihood that the Second Generation Attends College

	(1)	(2)	(3)	(4)	(5)
	OLS	IV			
	Attended College	Attended College	Attended College	Attended College	Attended College
TSP exposure ($10\mu g/m^3$)	0.00083**	−0.02050**	−0.01817*	−0.02152**	−0.02232**
	(0.00041)	(0.01010)	(0.00974)	(0.01051)	(0.01072)
Dep. Var. Mean	0.55	0.55	0.55	0.55	0.55
Observations	331,000	331,000	331,000	325,000	325,000
First Stage F-stat	–	23.82	23.83	23.7	23.45
1st Gen Fixed Effects	County-of-birth, State-of-birth × Year, Birth Month				
1st Gen Controls	Yes	Yes	Yes	Yes	Yes
2nd Gen Controls	No	No	Yes	Yes	Yes
2nd Gen County-of-birth FE?	No	No	No	Yes	Yes
2nd Gen State-of-birth×Year FE?	No	No	No	No	Yes

Notes: Significance levels are indicated as * 0.10 ** 0.05 *** 0.01. Source: Decennial Census Short Form 2000 and 2010, ACS 2005 through 2015. All regressions include the same set of control variables as in Table 22. Column 3-5 contain additional second generation demographic characteristics including race/ethnicity, sex and a quadratic in age. Standard errors are clustered at the parent's county of birth level. Approved for release by the Census DRB, authorization numbers CBDRB-FY18-232, CBDRB-FY18-248, CBDRB-FY18-287 and CBDRB-FY19-099.

Table 8: The Effect of In-Utero TSP Exposure on the Likelihood that the Second Generation Attends College, by Adopted Status

	(1)	(2)	(3)	(4)	(5)
	IV				
	ADOPTED	ATTENDED COLLEGE	ATTENDED COLLEGE	ATTENDED COLLEGE	ATTENDED COLLEGE
TSP exposure ($10 \mu g/m^3$)	-0.003	-0.02340^{**}	-0.02138^{**}	-0.02346^{**}	-0.02297^{**}
	(0.003)	(0.01081)	(0.01052)	(0.01108)	(0.01112)
TSP X Adopted		−0.00567	−0.01016	−0.00865	−0.01459
		(0.02305)	(0.02321)	(0.02343)	(0.02359)
1st Gen Fixed Effects	County-of-birth, State-of-birth × Year, Birth Month				
1st Gen Controls	YES	YES	YES	YES	YES
2nd Gen Controls	NO	NO	YES	YES	YES
2nd Gen County FE?	NO	NO	NO	YES	YES
2nd Gen SY FE?	NO	NO	NO	NO	YES
Observations	3,630,000	311,000	311,000	306,000	306,000
Control Mean	–	0.55	0.55	0.55	0.55
First Stage F-Stat	22.47	23.55	21.09	21.23	23.68

NOTES: Significance levels are indicated as * 0.10 ** 0.05 *** 0.01. Source: Decennial Census Short Form 2000 and 2010, ACS 2005 through 2015. All regressions include the same set of control variables as in Table 22. Column 3-5 contain additional second generation demographic characteristics including race/ethnicity, sex and a quadratic in age. Standard errors are clustered at the parent's county of birth level. Approved for release by the Census DRB, authorization numbers CBDRB-FY18-232, CBDRB-FY18-248, CBDRB-FY18-287 and CBDRB-FY19-099.

Table 9: The Effect of In-Utero TSP Exposure on Parental Time-Use

	(1)	(2)	(3)	(4)
	IV			
	Reading to Kids (minutes)	Reading to Kids (minutes)	Educational Activities (minutes)	Educational Activities (minutes)
TSP exposure ($10\mu g/m^3$)	-0.139^{**}	-0.146^{**}	-0.039	-0.053
	(0.068)	(0.069)	(0.106)	(0.106)
Fixed Effects	County-of-birth, State-of-birth × Year, Birth Month			
Individual Controls	Yes	Yes	Yes	Yes
County Controls	Yes	Yes	Yes	Yes
Day of Interview FE	No	Yes	No	Yes
Observations	9,000	9,000	9,000	9,000
Control Mean	1.44	1.44		
First Stage F-Stat	23.55	21.09	21.23	23.68

Notes: Significance levels are indicated as * 0.10 ** 0.05 *** 0.01. Source: Decennial Census Short Form 2000 and 2010, ACS 2005 through 2015. All regressions include individual demographic controls including sex, race/ethnicity and quadratic in age, pre-CAA 1970 county of birth economic characteristics (employment, total transfer income, personal income per capita) interacted with quadratic trends, county of birth weather controls including average and maximum temperature and number of precipitation days during an individual's 9 month gestational period. Standard errors are clustered at the parent's county of birth level. Approved for release by the Census DRB, authorization numbers CBDRB-FY18-232, CBDRB-FY18-248, CBDRB-FY18-287 and CBDRB-FY19-099.

Online Appendices – Not for Publication

Contents

A Robustness Checks

A.1 Reduced Form Results

Although we have provided some evidence consistent with the exclusion restriction holding, a key assumption underlying our IV strategy, this exclusion restriction is fundamentally untestable. However, in our setting, the reduced form effect of the nonattainment designations themselves is still of interest even if the exclusion restriction were in fact violated. We thus estimate regressions of the form:

$$Y_{j,c,s,t} = \alpha_c + \alpha_{s,t} + \beta_1 Nonattainment_{c,t} + X_j + e_{j,c,s,t}$$

for the first generation and

$$Y_{i,j,c,s,t} = \alpha_c + \alpha_{s,t} + \beta_1 Nonattainment_{c,t} + X_i + X_j + e_{i,j,c,s,t}$$

for the second generation.

For completeness, then, we produce the reduced form estimates of the effect of the CAAA 1970 on first and second generation outcomes in Tables A1 - A5. These results are all qualitatively similar to the IV results presented above, and largely have the same statistical properties. As would be expected given the properties of the IV estimator, the reduced form estimates are if anything slightly more precise than the IV estimates.

Table A1: Effect of CAAA 1970 Nonattainment Designations on First Generation Labor Market Outcomes

	(1)	(2)	(3)	(4)
	log Wages	log Wages (Below Median)	Unemployed	Public Assistance
Non-Attainment	0.00732*	0.00756*	-0.00001	0.00063
	(0.00419)	(0.0429)	(0.00069)	(0.00051)
Fixed Effects	County-of-birth, State-of-birth × Year, Birth Month			
Individual Controls	Yes	Yes	Yes	Yes
County-level Controls	Yes	Yes	Yes	Yes
Observations	3,391,000	1,475,000	4,766,000	4,767,000
Control Mean	$45,320	$25,870	0.048	0.020

Notes: Significance levels are indicated as * 0.10 ** 0.05 *** 0.01. Source: Decennial Census Short Form 2000 and 2010, ACS 2005 through 2015. All regressions include individual demographic controls including sex, race/ethnicity and quadratic in age, pre-CAA 1970 county of birth economic characteristics (employment, total transfer income, personal income per capita) interacted with quadratic trends, county of birth weather controls including average and maximum temperature and number of precipitation days during an individual's 9 month gestational period. Standard errors are clustered at the parent's county of birth level. Approved for release by the Census DRB, authorization numbers CBDRB-FY18-232, CBDRB-FY18-248, CBDRB-FY18-287 and CBDRB-FY19-099.

Table A2: Effect of CAAA 1970 Nonattainment Designations on First Generation Family Structure Outcomes

	(1) Divorced	(2) Married	(3) Any Kids	(4) Number of Kids	(5) Teen Pregnancy	(6) Age at Birth
Non-Attainment	-0.0022**	-0.0003	-0.0067	-0.0198*	0.0020*	0.0041
	(0.0009)	(0.0015)	(0.0046)	(0.0115)	(0.0012)	(0.0272)
Fixed Effects	County-of-birth, State-of-birth × Year, Birth Month					
Individual Controls	Yes	Yes	Yes	Yes	Yes	Yes
CountyControls	Yes	Yes	Yes	Yes	Yes	Yes
Observations	5,289,000	5,289,000	3,040,000	5,855,000	4,773,000	4,554,000
Control Mean	0.125	0.72	0.61	1.14	0.07	26

Notes: Significance levels are indicated as * 0.10 ** 0.05 *** 0.01. Source: Decennial Census Short Form 2000 and 2010, ACS 2005 through 2015. All regressions include individual demographic controls including sex, race/ethnicity and quadratic in age, pre-CAA 1970 county of birth economic characteristics (employment, total transfer income, personal income per capita) interacted with quadratic trends, county of birth weather controls including average and maximum temperature and number of precipitation days during an individual's 9 month gestational period. Standard errors are clustered at the parent's county of birth level. Approved for release by the Census DRB, authorization numbers CBDRB-FY18-232, CBDRB-FY18-248, CBDRB-FY18-287 and CBDRB-FY19-099.

Table A3: Effect of CAAA 1970 Nonattainment Designations on Second Gen Grade For Age

	(1)	(2)	(3)	(4)
	< Grade For Age	< Grade For Age	< Grade For Age	< Grade For Age
Non-Attainment	0.00123	0.00115	0.00132*	0.00077
	(0.00089)	(0.00080)	(0.00079)	(0.00076)
1st Gen Fixed Effects	County-of-birth, State-of-birth × Year, Birth Month			
1st Gen Controls	Yes	Yes	Yes	Yes
2nd Gen Controls	No	Yes	Yes	Yes
2nd Gen County FE?	No	No	Yes	Yes
2nd Gen SY FE?	No	No	No	Yes
Observations	2,626,000	2,626,000	2,582,000	2,582,000
Control Mean	0.28	0.28	0.28	0.28

Notes: Significance levels are indicated as * 0.10 ** 0.05 *** 0.01. Source: Decennial Census Short Form 2000 and 2010, ACS 2005 through 2015. All regressions include the same set of control variables as in Table 22. Column 3-5 contain additional second generation demographic characteristics including race/ethnicity, sex and a quadratic in age. Standard errors are clustered at the parent's county of birth level. Approved for release by the Census DRB, authorization numbers CBDRB-FY18-232, CBDRB-FY18-248, CBDRB-FY18-287 and CBDRB-FY19-099.

Table A4: Effect of CAAA 1970 Nonattainment Designations on Second Gen High School Noncompletion

	(1)	(2)	(3)	(4)
	High School Dropout	High School Dropout	High School Dropout	High School Dropout
Non-Attainment	0.00114	0.00048	0.00072	0.00070
	(0.00152)	(0.00151)	(0.00152)	(0.00151)
1st Gen Fixed Effects	County-of-birth, State-of-birth $\times$ Year, Birth Month			
1st Gen Controls	Yes	Yes	Yes	Yes
2nd Gen Controls	No	Yes	Yes	Yes
2nd Gen County FE?	No	No	Yes	Yes
2nd Gen SY FE?	No	No	No	Yes
Observations	1,305,000	1,305,000	1,281,000	1,281,000
Control Mean	0.06	0.06	0.06	0.06

Notes: Significance levels are indicated as * 0.10 ** 0.05 *** 0.01. Source: Decennial Census Short Form 2000 and 2010, ACS 2005 through 2015. All regressions include the same set of control variables as in Table 22. Column 3-5 contain additional second generation demographic characteristics including race/ethnicity, sex and a quadratic in age. Standard errors are clustered at the parent's county of birth level. Approved for release by the Census DRB, authorization numbers CBDRB-FY18-232, CBDRB-FY18-248, CBDRB-FY18-287 and CBDRB-FY19-099.

Table A5: Effect of CAAA 1970 Nonattainment Designations on Second Gen College Attendance

	(1)	(2)	(3)	(4)
	Attended College	Attended College	Attended College	Attended College
Non-Attainment	0.01668***	0.01553**	0.01875***	0.01870***
	(0.00638)	(0.00634)	(0.00643)	(0.00642)
1st Gen Fixed Effects	County-of-birth, State-of-birth × Year, Birth Month			
1st Gen Controls	Yes	Yes	Yes	Yes
2nd Gen Controls	No	Yes	Yes	Yes
2nd Gen County FE?	No	No	Yes	Yes
2nd Gen SY FE?	No	No	No	Yes
Observations	331,000	331,000	325,000	325,000
Control Mean	0.55	0.55	0.55	0.55

Notes: Significance levels are indicated as * 0.10 ** 0.05 *** 0.01. Source: Decennial Census Short Form 2000 and 2010, ACS 2005 through 2015. All regressions include the same set of control variables as in Table 22. Column 3-5 contain additional second generation demographic characteristics including race/ethnicity, sex and a quadratic in age. Standard errors are clustered at the parent's county of birth level. Approved for release by the Census DRB, authorization numbers CBDRB-FY18-232, CBDRB-FY18-248, CBDRB-FY18-287 and CBDRB-FY19-099.

A.2 Regression Discontinuity Design Approach

The baseline reduced form approach used throughout the paper treats all counties the same. However, it is likely that counties closer to the new air quality standards may be more affected by the nonattainment designations. As there is a sharp cutoff in assignment to nonattainment status, looking at the effects of nonattainment designations on outcomes in a regression discontinuity framework can shed additional light on the validity of our identification strategy.

The county-level ambient air quality standards set for particulate matter had two parts: annual average (geometric mean) TSP concentrations must be less than 75 $\mu g/m^3$, and the second highest daily average observation must be no less than 260 $\mu g/m^3$. In practice, the first part was binding for almost all counties – in our data, about 20 counties would have been in nonattainment due to the second part of the standard but not the first. We exclude these counties from the subsequent analysis and focus on the remaining counties.

Following Chay and Greenstone (2005) and Isen et al. (2017), we estimate parametric RDD regressions by supplementing our baseline reduced form estimates with a linear spline in the pre-CAA average TSP concentration. That is, we define $Dist_{c,t}$ as $TSP_c - 75$ (where TSP_c is the geometric mean TSP in 1970 in county c) for years after 1971, and set $Dist_c = 0$ for 1971 and earlier. We then estimate regressions of the form:

$$Y_{i,c,s,t} = \alpha_c + \alpha_{s,t} + \beta_1 Nonattainment_{c,t} + \beta_2 Dist_{c,t} + \beta_3 Dist_{c,t} \times I(TSP_c > 75) + \Gamma X_{i,c,s,t} + e_{i,c,s,t}$$

where $Y_{i,c,s,t}$ is a first or second gen outcome of interest. We estimate these RD regressions with bandwidths varying from 50 to 150 $\mu g/m^3$. As an additional check, we also restrict these regressions to a narrow window of parents born 1969-1974.

Table A6: Effect of CAAA 1970 Nonattainment Designations on First Generation Wages, RDD Approach

	Bandwidth (in $\mu g/m^3$)		
	50	100	150
Nonattainment	0.020**	0.009	0.009
	(0.008)	(0.008)	(0.008)
Fixed Effects	County, State × Year, Month		
Individual Controls	Yes	Yes	Yes
County-level Controls	Yes	Yes	Yes
Observations	714,000	1,029,000	1,061,000
Control Mean	$45,320	$45,320	$45,320

Notes: Significance levels are indicated as * 0.10 ** 0.05 *** 0.01. Source: Decennial Census Short Form 2000 and 2010, ACS 2005 through 2015. All regressions include individual demographic controls including sex, race/ethnicity and quadratic in age, pre-CAA 1970 county of birth economic characteristics (employment, total transfer income, personal income per capita) interacted with quadratic trends, county of birth weather controls including average and maximum temperature and number of precipitation days during an individual's 9 month gestational period. Standard errors are clustered at the parent's county of birth level.

Table A7: Effect of CAAA 1970 Nonattainment Designations on First Generation Divorce, RDD Approach

	Bandwidth (in $\mu g/m^3$)		
	50	100	150
Nonattainment	-0.005^{***}	-0.003^{*}	-0.003
	(0.002)	(0.002)	(0.002)
Fixed Effects	County, State × Year, Month		
Individual Controls	Yes	Yes	Yes
County-level Controls	Yes	Yes	Yes
Observations	714,000	1,029,000	1,061,000
Control Mean	$45,320	$45,320	$45,320

Notes: Significance levels are indicated as * 0.10 ** 0.05 *** 0.01. Source: Decennial Census Short Form 2000 and 2010, ACS 2005 through 2015. All regressions include individual demographic controls including sex, race/ethnicity and quadratic in age, pre-CAA 1970 county of birth economic characteristics (employment, total transfer income, personal income per capita) interacted with quadratic trends, county of birth weather controls including average and maximum temperature and number of precipitation days during an individual's 9 month gestational period. Standard errors are clustered at the parent's county of birth level.

Table A8: Effect of CAAA 1970 Nonattainment Designations on First Generation College Graduation, RDD Approach

	Bandwidth (in $\mu g/m^3$)		
	50	100	150
Nonattainment	0.007**	0.005*	0.006**
	(0.003)	(0.003)	(0.003)
Fixed Effects	County, State × Year, Month		
Individual Controls	Yes	Yes	Yes
County-level Controls	Yes	Yes	Yes
Observations	714,000	1,029,000	1,061,000
Control Mean	\$45,320	\$45,320	\$45,320

NOTES: Significance levels are indicated as * 0.10 ** 0.05 *** 0.01. Source: Decennial Census Short Form 2000 and 2010, ACS 2005 through 2015. All regressions include individual demographic controls including sex, race/ethnicity and quadratic in age, pre-CAA 1970 county of birth economic characteristics (employment, total transfer income, personal income per capita) interacted with quadratic trends, county of birth weather controls including average and maximum temperature and number of precipitation days during an individual's 9 month gestational period. Standard errors are clustered at the parent's county of birth level.

Table A9: Effect of CAAA 1970 Nonattainment Designations on Second Generation College Attendance, RDD Approach

	(1)	(2)	(3)	(4)
	Attended College	Attended College	Attended College	Attended College
Nonattainment	0.02514**	0.02206*	0.02534*	0.02790**
	(0.01273)	(0.01255)	(0.01306)	(0.01332)
1st Gen Fixed Effects	County-of-birth, State-of-birth × Year, Birth Month			
1st Gen Controls	Yes	Yes	Yes	Yes
2nd Gen Controls	No	Yes	Yes	Yes
2nd Gen County FE?	No	No	Yes	Yes
2nd Gen SY FE?	No	No	No	Yes
Observations	59,000	59,000	58,000	58,000
Control Mean	0.55	0.55	0.55	0.55

Notes: Significance levels are indicated as * 0.10 ** 0.05 *** 0.01. Source: Decennial Census Short Form 2000 and 2010, ACS 2005 through 2015. All regressions include the same set of control variables as in Table 22. Column 3-5 contain additional second generation demographic characteristics including race/ethnicity, sex and a quadratic in age. Standard errors are clustered at the parent's county of birth level.

Source: ACS 2005 through 2015, EPA Air Quality and Census Numident Data. Approved for release by the Census DRB, authorization numbers CBDRB-FY18-232, CBDRB-FY18-248, CBDRB-FY18-287 and CBDRB-FY19-099.

A.3 Alternate Sample Restrictions

We perform two additional robustness checks by varying the sample. First, we look at a narrow window of individuals born between 1969-1974 (this corresponds to the window in Isen et al. (2017)) an exercise similar in spirit to the RDD results above - by focusing on individuals born closer to the sharp decline in TSP that occurred in 1972, we may be more cleanly identifying effects, at the potential costs of lost precision.

Second, we vary the monitor sites used to measure county level TSP exposure. CAAA 1970 included not only additional regulatory authority, but also funding to, among other things, expand monitoring of ambient air pollution. Thus there was an expansion in the number of functioning monitors over the period 1969-1975. If these new monitors are placed endogenously to over or understate pollution levels, then using all available monitors may lead to endogenous measurement error, biasing the baseline results. To check this, we estimate our baseline first and second generation results using TSP levels calculated only using monitors active before 1972. Note that if the additional monitors result in measurement closer to the "true" level of pollution, then using these pre-1972 monitors will inject measurement error into our results, potentially attenuating our results and/or reducing precision of our estimates.

Table A10: Effect of CAAA 1970 Nonattainment Designations on Economic Outcomes, Pre-CAAA 1970 Monitors

	(1) Unemployed	(2) Public Assistance	(3) log(Wages)	(4) log(Wages) (Below Median)
Panel A: OLS estimates				
TSP exposure ($10\mu g/m^3$)	0.00004	0.00003	−0.00007	0.00008
	(0.00007)	(0.00005)	(0.00037)	(0.00048)
Panel B: IV estimates				
TSP exposure ($10\mu g/m^3$)	0.00079	−0.00068	−0.01868*	−0.02242*
	(0.00164)	(0.00113)	(0.01093)	(0.01183)
Fixed Effects	County-of-birth, State-of-birth × Year, Birth Month			
Individual Controls	Yes	Yes	Yes	Yes
County-level Controls	Yes	Yes	Yes	Yes
Control Mean	0.048	0.020	$45,320	$25,870
Observations	4,391,000	4,391,000	3,102,000	1,523,000
First Stage F-stat	21.52	21.53	21.87	23.47

Notes: Significance levels are indicated as * 0.10 ** 0.05 *** 0.01. Source: Census Numident, Decennial Census Short Form 2000 and 2010, ACS 2005 through 2015. All regressions include individual demographic controls including sex, race/ethnicity and quadratic in age, pre-CAA 1970 county of birth economic characteristics (employment, total transfer income, personal income per capita) interacted with quadratic trends, county of birth weather controls including average and maximum temperature and number of precipitation days during an individual's 9 month gestational period. Standard errors are clustered at the parent's county of birth level.

Source: ACS 2005 through 2015, EPA Air Quality and Census Numident Data. Approved for release by the Census DRB, authorization numbers CBDRB-FY18-232, CBDRB-FY18-248, CBDRB-FY18-287 and CBDRB-FY19-099.

Table A11: Effect of CAAA 1970 Nonattainment Designations on Family Outcomes, Pre-CAAA 1970 Monitors

	(1) Divorced	(2) Married	(3) Any Kids	(4) # Kids
Panel A: OLS estimates				
TSP exposure ($10\mu g/m^3$)	-0.0001	0.0003**	0.0002	0.0008
	(0.0001)	(0.0001)	(0.0002)	(0.0007)
Panel B: IV estimates				
TSP exposure ($10\mu g/m^3$)	0.0037	0.0010	0.0093	0.0360
	(0.0024)	(0.0033)	(0.0081)	(0.0223)
Fixed Effects	County-of-birth, State-of-birth × Year, Birth Month			
Individual Controls	Yes	Yes	Yes	Yes
County-level Controls	Yes	Yes	Yes	Yes
Control Mean	0.125	0.72	0.61	1.410
Observations	4,397,000	4,397,000	2,756,000	5,307,000
First Stage F	5.19	5.19	6.4	6.5

Notes: Significance levels are indicated as * 0.10 ** 0.05 *** 0.01. Source: Census Numident, Decennial Census Short Form 2000 and 2010, ACS 2005 through 2015. All regressions include individual demographic controls including sex, race/ethnicity and quadratic in age, pre-CAA 1970 county of birth economic characteristics (employment, total transfer income, personal income per capita) interacted with quadratic trends, county of birth weather controls including average and maximum temperature and number of precipitation days during an individual's 9 month gestational period. Standard errors are clustered at the parent's county of birth level.

Source: ACS 2005 through 2015, EPA Air Quality and Census Numident Data. Approved for release by the Census DRB, authorization numbers CBDRB-FY18-232, CBDRB-FY18-248, CBDRB-FY18-287 and CBDRB-FY19-099.

Table A12: Effect of TSP Exposure on Second Generation College Attendance, Pre-CAAA 1970 Monitors

	(1)	(2)	(3)	(4)	(5)
	OLS	IV			
	Attended College	Attended College	Attended College	Attended College	Attended College
TSP exposure ($10\mu g/m^3$)	0.00101**	−0.05074	−0.04331	−0.05109	−0.05243
	(0.00040)	(0.04155)	(0.03715)	(0.04301)	(0.04469)
Dep. Var. Mean	0.55	0.55	0.55	0.55	0.55
Observations	327,000	327,000	327,000	321,000	321,000
First Stage F-stat	–	23.82	23.83	23.7	23.45
1st Gen Fixed Effects	County-of-birth, State-of-birth × Year, Birth Month				
1st Gen Controls	Yes	Yes	Yes	Yes	Yes
2nd Gen Controls	No	No	Yes	Yes	Yes
2nd Gen County-of-birth FE?	No	No	No	Yes	Yes
2nd Gen State-of-birth×Year FE?	No	No	No	No	Yes

Notes: Significance levels are indicated as * 0.10 ** 0.05 *** 0.01. Source: Decennial Census Short Form 2000 and 2010, ACS 2005 through 2015. All regressions include the same set of control variables as in Table 22. Column 3-5 contain additional second generation demographic characteristics including race/ethnicity, sex and a quadratic in age. Standard errors are clustered at the parent's county of birth level.

Source: ACS 2005 through 2015, EPA Air Quality and Census Numident Data. Approved for release by the Census DRB, authorization numbers CBDRB-FY18-232, CBDRB-FY18-248, CBDRB-FY18-287 and CBDRB-FY19-099.

Table A13: Effect of TSP Exposure on Economic Outcomes, Parents Born 1969-1974

	(1) UNEMPLOYED	(2) PUBLIC ASSISTANCE	(3) log(WAGES)	(4) log(WAGES) (Below Median)
Panel A: OLS estimates				
TSP exposure ($10\mu g/m^3$)	0.00011	0.00007	−0.00054	−0.00074
	(0.00014)	(0.00009)	(0.00065)	(0.00090)
Panel B: IV estimates				
TSP exposure ($10\mu g/m^3$)	0.00050	0.00060	−0.01028*	−0.00837
	(0.00100)	(0.00064)	(0.00557)	(0.00741)
Fixed Effects	County-of-birth, State-of-birth × Year, Birth Month			
Individual Controls	YES	YES	YES	YES
County-level Controls	YES	YES	YES	YES
Control Mean	0.048	0.020	$45,320	$25,870
Observations	1,567,000	1,567,000	1,117,000	556,000
First Stage F	17.2	17.21	17.87	19.17

NOTES: Significance levels are indicated as * 0.10 ** 0.05 *** 0.01. Source: Census Numident, Decennial Census Short Form 2000 and 2010, ACS 2005 through 2015. All regressions include individual demographic controls including sex, race/ethnicity and quadratic in age, pre-CAA 1970 county of birth economic characteristics (employment, total transfer income, personal income per capita) interacted with quadratic trends, county of birth weather controls including average and maximum temperature and number of precipitation days during an individual's 9 month gestational period. Standard errors are clustered at the parent's county of birth level.

Source: ACS 2005 through 2015, EPA Air Quality and Census Numident Data. Approved for release by the Census DRB, authorization numbers CBDRB-FY18-232, CBDRB-FY18-248, CBDRB-FY18-287 and CBDRB-FY19-099.

Table A14: Effect of TSP Exposure on Family Outcomes, Parents Born 1969-1974

	(1) Divorced	(2) Married	(3) Any Kids	(4) # Kids
Panel A: OLS estimates				
TSP exposure ($10\mu g/m^3$)	0.00004	−0.00029	−0.00028	−0.00004
	(0.00017)	(0.00022)	(0.00030)	(0.00074)
Panel B: IV estimates				
TSP exposure ($10\mu g/m^3$)	0.00258	−0.00024	0.00038	0.00820
	(0.00167)	(0.00191)	(0.00284)	(0.00661)
Fixed Effects	County-of-birth, State-of-birth × Year, Birth Month			
Individual Controls	Yes	Yes	Yes	Yes
County-level Controls	Yes	Yes	Yes	Yes
Control Mean	0.125	0.72	0.61	1.410
Observations	1,570,000	1,570,000	949,000	1,834,000
First Stage F	17.22	17.22	16.61	16.88

Notes: Significance levels are indicated as * 0.10 ** 0.05 *** 0.01. Source: Census Numident, Decennial Census Short Form 2000 and 2010, ACS 2005 through 2015. All regressions include individual demographic controls including sex, race/ethnicity and quadratic in age, pre-CAA 1970 county of birth economic characteristics (employment, total transfer income, personal income per capita) interacted with quadratic trends, county of birth weather controls including average and maximum temperature and number of precipitation days during an individual's 9 month gestational period. Standard errors are clustered at the parent's county of birth level.

Source: ACS 2005 through 2015, EPA Air Quality and Census Numident Data. Approved for release by the Census DRB, authorization numbers CBDRB-FY18-232, CBDRB-FY18-248, CBDRB-FY18-287 and CBDRB-FY19-099.

Table A15: Effect of TSP Exposure on Second Generation College Attendance, Parents Born 1969-1974

	(1)	(2)	(3)	(4)	(5)
	OLS	IV			
	Attended College	Attended College	Attended College	Attended College	Attended College
TSP exposure ($10\mu g/m^3$) 0.00017	-0.01447	-0.01212	-0.01550	-0.01527	
	(0.00974)	(0.00974)	(0.00945)	(0.01017)	(0.01032)
Dep. Var. Mean	0.55	0.55	0.55	0.55	0.55
Observations	93,500	93,500	93,500	92,000	92,000
First Stage F-stat	– 12.72	12.7	12.97	13	
1st Gen Fixed Effects	County-of-birth, State-of-birth × Year, Birth Month				
1st Gen Controls	Yes	Yes	Yes	Yes	Yes
2nd Gen Controls	No	No	Yes	Yes	Yes
2nd Gen County-of-birth FE?	No	No	No	Yes	Yes
2nd Gen State-of-birth×Year FE?	No	No	No	No	Yes

Notes: Significance levels are indicated as * 0.10 ** 0.05 *** 0.01. Source: Decennial Census Short Form 2000 and 2010, ACS 2005 through 2015. All regressions include the same set of control variables as in Table 22. Column 3-5 contain additional second generation demographic characteristics including race/ethnicity, sex and a quadratic in age. Standard errors are clustered at the parent's county of birth level.

Source: ACS 2005 through 2015, EPA Air Quality and Census Numident Data. Approved for release by the Census DRB, authorization numbers CBDRB-FY18-232, CBDRB-FY18-248, CBDRB-FY18-287 and CBDRB-FY19-099.

B Data Appendix

B.1 Census Data Linkage Infrastructure

Note: this section describes the overall Census data linkage infrastructure. The procedures described here have already been performed on the analysis data we work with. This appendix describes the use of Title 13 Census data to identify parent-child links. Authorization from the IRS to use additional parent-child links identified with tax data is in process; upon receiving approval the results above will be revised to utilize these new links.

The U.S. Census Bureau is authorized, under Titles 13 and 26 of the US Code, to utilize all available data resources, including administrative records and commercially provided data, to improve the measurement of the US population and economy. Under this authority, the Census Bureau has developed a data linkage infrastructure which allows researchers to integrate data from multiple sources, including administrative records from federal and state government agencies, Decennial Census data, and demographic surveys. The central component of the Data linkage infrastructure is the Person Identification Validation System (PVS), which is described in further detail in Wagner and Layne (2014).

PVS is designed as a flexible probabilistic matching system that can be deployed in production to analyze very large datasets in a computationally efficient manner. PVS has two components: a person-based matching algorithm and an address-based matching algorithm. The address based matching algorithm takes a string address as an input (e.g. "1600 Pennsylvania Ave NW, Washington, DC 20001"), splits the string into components (street number, street name, street suffix, city, state, zipcode), standardizes these components, and then matches the address to a reference file (the Census Master Address File), optimizing on a fuzzy string comparator (the Levanstein string distance). The person based matching algorithm has a similar structure: it takes as input the available personally identifiable information on a file (name, SSN, date of birth, sex, address), and, after standardization, matches these PII fields to a separate reference file (the Census Numident).

Each of these matching algorithms produces a unique anonymized identifier for each successful match. For the address matching algorithm, the resulting identifier is called a MAFID (Master Address File Identifier), while the person-based matching algorithm uses PIKs (protected identification keys). MAFIDs and PIKs are both static hashes referencing a single entity in the relevant reference file, and can thus be used to link datasets without including any personally identifiable information on the research files used by researchers. Any attempt to infer PII from a research file with PIKs or MAFIDs is thus a violation of Title 13, with potential punishments including 10 years in prison, and hundreds of thousands of dollars in fines.

Not all of the PII inputs used by the PVS system are found in every microdata file on which PVS is applied. In particular, Social Security numbers are rarely elicited on demographic surveys, and have never been asked for in decennial Censuses.[17] Administrative

[17]The Current Population Survey ASEC asked for SSNs until 2002; however, non-response increased dramatically through the 1990s. This was in fact one of the motivating factors in the development of PVS. Moving from SSN-based matching to PVS-based probabilistic matching actually increased match rates for the CPS after 2002.

records which contain SSNs (e.g. most tax records) can be assigned a PIK in 99+ percent of cases. Match rates are still high for many demographic surveys and the decennial census, which ask for name and exact date of birth. The PIK assignment rate for the 2010 Census is about 91 percent, while the PIK assignment rate for the 2013 American Community Survey is about 94 percent.

B.2 Parent-Child Links

To study the intergenerational effects of the Clean Air Act, it is necessary to locate parents at birth (around the enactment of CAA 1970), link these parents to their children, and measure outcomes for both parents and children. We begin by assembling a database of all parent-child links that can be evaluated using the various data sources available in the Census Data Linkage Infrastructure. The set of links we are able to identify is not, we should stress, the full population of links. In our empirical analysis, we will attempt to re-weight the data to address the fact that the missing links we are not able to identify are almost certainly not missing at random.

To benchmark our link coverage, consider that the completed cohort fertility rate for women born in 1970 is about 2.1. There were about 44 million women aged 30-50 in the 2010 Census (i.e. born between 1960-1980). Taking the 1970 CCFR as constant throughout this group, we can expect at most 92 million natural born children. In practice we will identify fewer than this, due to linkage error, and the fact that women born in the latter part of our birth year range will not have completed fertility in the latest available data we are using to identify parent-child links (the 2015 ACS).

B.2.1 Decennial Census Data

The 2000 and 2010 decennial Census 100 percent detail file (HDF), colloquially the "Census short form", collects an abbreviated set of demographic information from the full population of the United States in decadal Census years. This demographic information includes date of birth, sex, race and ethnicity, and some relationship information. Unfortunately, the relationship information collected in the Census does not capture the full relationship structure within a household. Rather, the Census collects information from each individual in a household on their relationship to the primary household member (the first person listed on the census form for the household), coded as the variable QREL.

This means it is possible to identify two types of parent child links: "certain" parent-child links between a child and the householder parent, and "probable" parent-child links between a child and the married or unmarried partner of their parent householder. The relationship codes are sufficiently detailed to separate natural born children of a householder (QREL code 3), adopted children (QREL code 4) and stepchildren (QREL code 5). For the purposes of the project at hand, we identify only parent-child links (certain or probable) for parents born between 1960-1982.

To identify these two types of links in the 2000 Census HDF, we use the following algorithm. We first subset the HDF by age and relationship code, retaining only individuals aged 40 or younger (i.e. who were born after 1960) who have QREL codes 1 (householder), 2 (spouse of householder), 3 (natural born child of householder), 4 (adopted child of householder, 5 (stepchild of householder) or 19 (unmarried partner of householder). Then, for each household, we assign three link variables: "Certain Parent", which is the PIK of the householder, "Probable-Married", which is the PIK of the householder's spouse, and "Probable-Unmarried", which is the PIK of the householder's unmarried partner. Each of these variables are missing if the relevant PIK is missing (due to PIK non-assignment when the HDF was analyzed via PVS). We then reshape the data into long form (so each row contains the child's PIK, the parent's PIK and indicators for the type of child and the type of parent). We discard all cases where the child or parent's PIKs are missing.

This yields a dataset containing about 65 million parent-child links. Of these, about 35 million are "Certain" Links", about 28 million are "Probable-Married", and the remaining approximately 2 million are "Probable-Unmarried". We identify more mother-child links ($\approx$ 38 million) than father-child links (27 million). As expected, the parent-child links identified in the 2000 HDF are heavily tilted toward the older parents: about 51 million links involve parents born before 1970, while about 14 million involve parents born after 1980.

We repeat the use of the same algorithm to identify parent-child links in the 2010 HDF. We identify substantially more links in the 2010 Census, as expected. In all, we identify 115 million parent-child links – of these, about 64 million are "Certain", 46 million are "Probable-Married" and the remaining 5 million are "Probable-Unmarried". As with the 2000 HDF, we identify more mother-child links (65 million) compared to father-child links (50 million). We continue to identify more parent-child links for parents born before 1970, although the split is much more even compared to the 2000 HDF (reflecting the fact that women born before 1970 had largely completed fertility, while women born after 1970 were still in prime childbearing age ranges).

Combining the information from the two decennial Census files, we can identify about 152 million unique parent-child links for about 81 million children. Note that because of the way that the "Probable" links are identified, it is possible that some of these links represent changes in family structure (marriages, divorces, and creation/dissolution of unmarried partnerships). About 123 million links occur for children with 1 or 2 unique links, while the remaining 29 million occur for children with 3 or more links (these represent cases where parental relationships appear to have changed).

B.2.2 Other Demographic Surveys

The final source of data on parent-child links comes from demographic surveys. These surveys are substantially smaller than the Census, but allow us to identify relationships in non-Decennial year. We use two such surveys: the Current Population Survey Annual Social and Economic Supplement (CPS ASEC), which is conducted every March, and the American Community Survey, which has been conducted monthly since 2001. The CPS ASEC sample size is substantially smaller than the ACS (200,000 individuals in the CPS-ASEC versus about 5 million in the ACS 1 year files), however, the CPS ASEC is available for a longer period of time – CPS ASEC has been conducted annually since 1968, although only the ASEC files after 1990 have been assigned PIKs.

The ACS was conducted as an experimental survey from 2001-2004, with increasing sample size in each year. From 2005–2015, the ACS has consisted of a sample size of about 5 million individuals. The content of the ACS has evolved considerably over this period.

In particular, the household relationship question was redesigned for the 2008 ACS. From 2001–2007, the ACS relationship question allowed for 10 categories, with a single "child of householder" category that includes adopted children, natural born children and step-children. From 2008–2015, the relationship question was expanded to 13 categories, with separate categories for adopted children, natural born children and step-children. As with the Decennial HDF data, the relationship variable in the ACS contains categories for married and unmarried partners of the householder, so we categorize the ACS links as "Certain" (for the householder), "Probable-Married" (for the householder's spouse) and "Probable-Unmarried" (for the householder's unmarried partner). We are able to collect about 22 million parent-child links for about 12 million children.

B.2.3 Combining Relationship Information

The relationship information we have extracted from Decennial Census data and demographic surveys has a substantial degree of overlap. In total, we identify links for over 168 million children.

Note while a vast majority (about 87 percent) of children can be linked to one or two parents, there are a substantial number who are linked to three or more parents.

We initially retain links from each source, to allow for robustness checks on the type of link used (i.e. just using Census links or keeping only "certain" parent-child links). Some source information is included in the data, including the parent and child types from the Census and survey data and the year(s) a link appears in the ACS data. Since the Census data sources are associated with specific years (2000 and 2010), it is also possible to select links based on vintage, e.g. selecting the first set of parent-child links which occur in the data (these are more likely to be biological parents).

B.3 Pollution Exposure at Birth

To analyze the intergenerational effects of pollution exposure, we need to be able to infer the level of ambient air pollution and the changes in EPA policy (designation nonattainment of NAAQS) that parents were exposed to at birth. We do this in three steps. First, we link the set of unique parents identified in the previous section to the Census Numident to obtain date and place of birth. We then obtain monitor-level daily pollution measures from the EPA, which aggregate to the county level, and link these county-level measures to the parents' place of birth. Finally, since the EPA's records of nonattainment designations appears to be incomplete or destroyed, we simulate these nonattainment designations for counties with EPA monitors active in 1969 (before CAA 1970).

B.3.1 Census Numident Data

Our source of information on the parents' place of birth comes from the Census Numident, which is a derivative product of the SSA Numerical Identification File, and serves as the reference file for the PVS matching algorithm. The Census Numident contains three fields which can be used to infer place of birth, which are transcribed from form SS-5 (application for social security number). The field *pobfin* contains a two digit code for the country of birth for non-native born individuals, and the field *pobst* contains a two character abbreviation for state of birth for all native born US citizens. Both of these fields can be assigned one-to-one with standard geographies (i.e. FIPS codes). The field *pobcity*, however, is slightly more cumbersome. This variable represents the first 12 characters of the place (or county) of birth entered on form SS-5. There is little standardization or cleaning done by SSA or Census for this field, and thus there are numerous misspellings and inconsistencies.

In order to match the information in the *pobcity* with standardized geographies (i.e. county FIPS codes), we take a two-step approach. First, after excluding foreign-born individuals (about 13 million parents), we capitalize on a crosswalk developed jointly by Census researchers and external researchers including Martha Bailey and Reed Walker. This crosswalk provides all exact matches (after standardization) and probabilistic matches between *pobcity* entries and unique GNIS place names. A second crosswalk between GNIS places and county FIPS codes allows us to directly match parents to counties exactly. For the remaining cases, we execute a probabilistic matching algorithm. This algorithm assigns a match by calculating the optimal string alignment (OSA) distance between a *pobcity* entry and a reference list of all county and Census place names, selecting the smallest distance (maximum of 5) within *pobst*. This is essentially the same algorithm as in Voorheis (2017a). All told, about 74 percent of native-born parents can be assigned a place of birth using the GNIS crosswalks, and another 23 percent can be matched using our probabilistic matching algorithm, so that about 97 percent of native born parents can be assigned a county of birth.

B.3.2 EPA Monitor Data

With information about the place of parents' birth in hand, we infer the level of pollution exposure experienced by these individuals if we have some information based on the average exposure within their county of birth. To gather this pollution exposure information, we rely on monitor data from the EPA. The EPA has made monitor-level air quality data available via the AQDM API. Our pollutant of interest is particulate matter. For the relevant period of time (around 1970), the primary regulated pollutant was total suspended particulates (TSP), defined as the density of particulates less than 50 microns, measured in units $\mu g/m^3$.[18] We thus retrieve all TSP (EPA pollutant code 11101) monitor observations between 1960–1980.

The TSP standard was set based on a 24-hour sampling, and hence the monitor-level data is provided on a daily basis. Our baseline approach to aggregating these daily monitor-level observations is as follows. For each county-day, we calculate the average TSP concentration across all active monitors in that day, which we take as the average exposure to TSP in that county on that day. We then calculate county-level moving average exposure to TSP for each unique birthday between 1960 and 1980 for two periods of interest: the nine months before birth (in utero exposure) and the year after birth (infant exposure).

The EPA's monitoring network expanded dramatically following the passage of CAA 1970, expanding both the number of counties monitored and the density of monitors within consistently monitored counties. This poses two potential challenges to our baseline measurement approach above. First, some counties will only have observations in the "post-treatment" period in our OLS and IV regressions. Second, even for counties which are

[18] This definition was later revised to 10 microns (PM10) and 2.5 microns (PM2.5) standards in 1987 and 1997 respectively.

consistently monitored, the expansion of the monitor network may result in systematic measurement error – average county exposure will be more precisely measured with more monitors and so the pre-treatment observations are more likely to be mismeasured than the post-treatment observations. To address these issues, we also produce county-level moving averages using a constant set of monitors (the monitors that were active in 1969 or earlier).

B.3.3 Nonattainment Designations

Our empirical strategy relies identifying the intergenerational effects of pollution exposure at birth using plausibly exogenous variation in TSP exposure that resulted from counties being designated as in nonattainment of the ambient air quality standards in the CAA 1970 by the EPA. Although the EPA makes nonattainment designations publicly available starting in 1991, and researchers have reconstructed nonattainment designations back to 1980, there appear to be no existing records on which counties were initially designated as being in nonattainment in 1972, the first year in which the CAA 1970 was in effect. The TSP air quality standards are known, however, and as noted in the previous section, we have monitor-level data on the actual level of exposure in the years before the CAA 1970 was in effect. Thus it is possible to reconstruct which counties would have been designated as in non-attainment.

Nonattainment of the primary air quality standard for TSP set in CAA 1970 occurs if either a) a county's annual average (geometric mean) TSP concentration is above 75 $\mu g/m^3$, or b) the second highest daily TSP concentration is above 260 $\mu g/m^3$. We use the monitor-level observations from the previous section to calculate the geometric mean and second highest daily TSP concentration for all counties with at least on monitor in 1970. This allows us to categorize 258 counties as "nonattainment" counties, and 319 counties as "attainment" counties.

B.3.4 Other County Attributes

Estimating the effects of pollution exposure at birth on adult outcomes for parents and intergenerational effects for their children may be confounded by other characteristics of the parents' place of birth, such as weather or economic activity. To this end, we obtain pre-determined (i.e. before the clean air act of 1970) information on county level economic activity from the BEA, and county-level weather information from NOAA.

Following Isen, et al. (2017), we obtain information on the economy and population of U.S. counties in 1969 from the Bureau of Economic Analysis' Regional Economic Accounts (1969 is the earliest year for which the BEA publishes regional accounts data). We extract four variables of interest from the regional accounts: total population, total employment, total personal income and total personal transfer income. From these we can construct income per capita and employment-to-population ratio measures; these measures allow us to control for important county-level economic characteristics that may confound the nonattainment-pollution relationship.

Additionally, we obtain information on county-level weather patterns. Temperature and precipitation, in particular, play important factors in the formation of particulate matter emissions and in the suspension of particulate matter in the atmosphere after emission. Importantly, there is evidence both that very low temperatures can increase PM concentrations by emissions (at cold temperatures, internal combustion engines burn fuel less efficiently), while very high temperatures can increase PM concentrations through suspension and atmospheric particle formation (sulfate and nitrate particles form more readily at hot temperatures. Additionally, precipitation decreases PM concentrations by decreasing suspension. Thus we obtain weather-station level data on daily high temperature, low temperature and precipitation from NOAA's Global Historical Climatology Network (GHCN). For each day between 1959 and 1981, we interpolate across the weather station network to each county centroid using inverse distance weighting to obtain a county-day level dataset. We can then calculate the average high/low temperature and number of precipitation days corresponding to the 9 months before birth and the year after birth for each individual.

B.4 ATUS Data

To investigate mechanisms underlying the second generation effect, we will leverage a secondary linked dataset which will allow us to measure both time use for individuals at a point in time, as well as their place of birth and the level of pollution they were exposed to. We do this by linking a subset of respondents to the American Time Use Survey (ATUS) to the Census Numident.

Using the IPUMS public use ATUS data from 2003-2017, we build a series of time use variables which divide the total time spent during the reference day on specific child-enrichment activities (time spent on children's education activities, time spent on children's health activities, time spent reading to a child), as well as broad categories of non-sleep time use (time spent on work, time spent on social activities, time spent on leisure, time spent on education). We then link a subset of the ATUS respondents to the Census Numident to attach place of birth characteristics as follows.

Our linkage strategy relies on the fact that the ATUS sample frame is drawn from the Current Population Survey. Hence it is possible to link ATUS respondents to the CPS on an individual level in the public use data. For the subset of individuals who are in sample and respond to the ASEC, we can link this public use identifier to the internal confidential CPS-ASEC data. The internal CPS-ASEC has had PIKs assigned, so we are then able to link these subset of individuals to the Census Numident by PIK, identifying place of birth and TSP exposure at birth using the same method used for the ACS sample, described above. We further subset this linked sample to individuals born 1960-1980, coinciding with the first generation for the main ACS results. Note that this is a relatively small subsample of ATUS respondents (the final analysis sample has about 10,000 observations).