Washington Center for Equitable Growth
Evidence for a stronger economy

equitablegrowth.org

Working paper series

The Grandkids Aren't Alright: The Intergenerational Effects of Prenatal Pollution Exposure

Jonathan Colmer
John Voorheis

February 2021

https://equitablegrowth.org/working-papers/the-grandkids-arent-alright-the-intergenerational-effects-of-prenatal-pollution-exposure/

The Grandkids Aren't Alright: The Intergenerational Effects of Prenatal Pollution Exposure.*

Jonathan Colmer
University of Virginia

John Voorheis
U.S. Census Bureau

Abstract

Evidence increasingly shows that environmental quality shapes human capital at birth with long-run effects on health and welfare. Do these effects, in turn, affect the economic opportunities of future generations? Using newly linked survey and administrative data, providing more than 150 million parent-child links, we show that regulation-induced improvements in air quality that parents experienced during gestation increase the likelihood that their children, the second-generation, attend college 40-50 years later. Greater parental resources and investments, rather than biological channels, appear to drive this effect. Our findings suggest that within-generation estimates of marginal damages substantially underestimate the total welfare effects of improving environmental quality and point to the empirical relevance of environmental quality as a contributor to economic opportunity in the United States.

*This Version: November 2020. 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, Alan Barecca, Andrew Barr, Raj Chetty, Janet Currie, Chloe East, James Fenske, Leora Friedberg, John Friedman, Chloe Gibbs, Michael Greenstone, Nathaniel Hendren, Bill Johnson, Nicolai Kuminoff, Jason Lindo, Marianne Page, Maya Rossin-Slater, Chris Ruhm, Nicholas Sanders, Jay Shimshack, Bill Shobe, Jeffrey Shrader, Sandip Sukhtankar, Sarah Turner, 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, UC Berkeley, the University of Birmingham, Cambridge, Columbia, the University of Exeter, Georgetown University, Harvard University, the Heartland Environmental & Resource Workshop at Illinois, Imperial College London, the London School of Economics, the University of Oregon, the University of Oxford, the NBER Children's Spring Meeting, the University of Virginia, the U.S. Census Bureau, and the University of Warwick 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, CBDRB-FY2020-CES010-001, CBDRB-FY2020-CES010-015, CBDRB-FY2020-CES010-017, CBDRB-FY2020-CES010-029, CBDRB-FY2020-CES010-031, CBDRB-FY2020-CES010-001 and CBDRB-FY21-CES014-004

1 Introduction

Growing evidence suggests that gestational exposure to air pollution, and other environmental factors, plays an important role in shaping endowments at birth, with long-term effects on health and welfare (Chay and Greenstone, 2003a,b; Almond, 2006; Black et al., 2007; Currie et al., 2009; Kelly, 2011; Almond et al., 2010; Isen et al., 2017; Black et al., 2018). A separate literature has considered the importance of parental income and human capital for the economic opportunities of their children (Becker and Tomes, 1979; Cunha and Heckman, 2007; Heckman and Mosso, 2014; Lee and Seshadri, 2019) and documented high levels of intergenerational persistence (Solon, 1992; Black et al., 2005; Black and Devereux, 2011; Chetty et al., 2014; Chetty and Hendren, 2018a,b; Chetty et al., 2018). Taken together these findings raise an important question: Do environmental shocks to endowments at birth have persistent effects across generations? The answer to this question has important implications for determining the returns to investments in environmental quality, as well as for understanding the contribution that environmental quality plays in shaping economic opportunity and social mobility.

We provide evidence that prenatal particulate matter exposure has intergenerational consequences. To identify this effect 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 particles (TSP), resulting in substantial improvements in ambient air quality in hundred of counties across the United States. This variation has been previously used to explore the effects of air pollution exposure on contemporaneous outcomes such as infant mortality (Chay and Greenstone, 2003a), home prices (Chay and Greenstone, 2005), fetal mortality (Sanders and Stoecker, 2015), adult mortality (Chay et al., 2003), as well as later-life earnings for those that were born at the time of the 1970 CAAA (Isen et al., 2017). For good reason the existing literature focuses on those that were directly affected by the 1970 CAAA. However, if regulation-induced reductions in pollution exposure have intergenerational consequences then within-generation analyses could

substantially underestimate the total welfare effects of investments in environmental quality.

We focus on the children of those who directly benefited from the 1970 CAAA. We compare outcomes for the children of cohorts who were born just before these changes went into effect to the children of cohorts that were born just after these large changes in air quality. This isolates the additional effect of parental exposure to clean air in very early childhood relative to improvements in the air quality of parents born before the introduction of the 1970 CAAA. The key identification assumption is that all cohorts experienced the same air pollution later in life. In support of this assumption we estimate that, on average, there are no differences in later life particulate matter exposure for the first-generation, nor any differences in prenatal or later-life particulate matter exposure for the second-generation.

We are able to explore the intergenerational consequences of improvements in environmental quality thanks to a new set of parent-child linkages constructed using U.S. Census Bureau data and administrative records. Our data exploits the exact date and location of birth for all children born in the 1960s and 1970s and more than 150 million parent-child links.

Our results suggest that prenatal exposure to particulate matter has a statistically significant and economically meaningful impact not only on those directly affected but also on the economic opportunities of their children. We estimate that children whose parents experienced lower gestational pollution exposure are much more likely to attend college. A $10\mu g/m^3$ reduction in parental gestational TSP exposure is associated with a 3.2-3.9 percentage point increase in the likelihood of attending college.[1] This is equivalent in magnitude to the intergenerational effects of 0.2-0.25 Head Start Programs (Barr and Gibbs, 2017). Back-of-the-envelope calculations suggest that the economic consequences of the estimated college enrollment effect are meaningful. Assuming a college wage premium of $28,000 p.a.

[1]The implied estimate from the reduced form effect of the 1970 CAAA, which, on average, induced an $8\mu g/m^3$ reduction for the parents of the second-generation sample, is almost identical. An 8 $\mu g/m^3$ reduction in parental gestational TSP exposure is associated with a 2.6-3.08 percentage point increase in the likelihood of attending college. Our reduced form estimates range between a 0.26-3 percentage point effect size. In proportional terms, a 10 $\mu g/m^3$ decrease in TSP exposure from the 1971 baseline would be similar to the cross-sectional difference in particulate matter exposure between blacks and whites in the 2000's.

(March CPS), a graduation rate of 50 percent (National Center for Education Statistics), a 3 percent real discount rate, and a constant earnings effect over the lifecycle, we calculate that a $10\mu g/m^3$ reduction in parental gestational TSP exposure amounts to a \$1,515 increase in cumulative lifetime earnings in present value terms. Combining the number of births in nonattainment counties in 1972 (1.52 million), the probability that an individual had a child (0.63), and the average number of children born to each individual (2.24), we calculate that the total second generation earning effects associated with TSP reductions for the 1972 cohort are ≈\$3.2 billion. To the extent that the estimated reductions in TSP have been permanent these benefits continue to accrue for the children of all cohorts born since. To the degree that the intergenerational effects of pollution exposure have effects on income other than through college our estimates represent a lower-bound.

In light of these effects we investigate the mechanisms that drive intergenerational transmission. We posit that there are two broad potential channels. The first channel is biological. Prenatal pollution exposure could result in epigenetic changes, i.e, permanent changes in gene expression, that are hereditarily transmitted to the second-generation. The second mechanism is economic. Prenatal pollution exposure affects the health, human capital, and earnings of the first generation. In turn, parental resources and investments may affect the likelihood of college attendance for their children.

First, we explore the empirical relevance of the biological pathway. We do so using information on the adopted status of children. We find no meaningful differential effect between adopted and biological children.[2] This finding suggests that the college attendance effects are unlikely to be driven by the heritable transmission of any epigenetic changes that were realized by the first generation. We note that this does not imply that heritable transmissions did not occur. There may be latent health effects associated with inherited epigenetic changes that are not realized at this stage of the life cycle.

Second, we explore whether our estimated effects might be driven by economic forces

[2]We do not find any effect of prenatal pollution exposure on the likelihood of the first-generation adopting a child. This suggests that there is not differential selection into adoption by treatment status.

— parental resources and investments. We estimate that the regulation-induced reductions in prenatal TSP exposure resulted in a 1.2 percent increase in mean annual earnings for affected cohorts — a $3,499 increase in cumulative lifetime earnings in present value terms.[3] While this effect is non-trivial, the existing literature suggests that the effect size is likely too small to plausibly explain the entirety of our college-attendance estimate (Lovenheim, 2011; Lovenheim and Lockwood Reynolds, 2013; Bulman et al., 2017). However, the existing literature explores the effects of an increase in individual wealth much later in life (usually during the teenage years of the second generation). Our earnings effects are experienced by the whole community, and because the first-generation were affected at birth, the second-generation benefit from these effects for their entire life, including the early years where increased household resources may be particularly important (Bastian and Michelmore 2018, Bailey et al. 2020). Nevertheless, it is likely that other factors are important as well.

We also explore the degree to which differential selection into our second-generation sample could drive our effects, a mechanism at the intersection of biology and economics. Based on existing evidence the cohorts that experienced lower prenatal particulate matter exposure end up healthier and richer than cohorts that were born prior to the introduction of the 1970 CAAA. This is not a threat to identification as the fundamental cause for these differences is still prenatal pollution exposure. However, it is possible that income and health affect fertility patterns (Becker, 1960; Lovenheim and Mumford, 2013; Black et al., 2013; Dettling and Kearney, 2014; Kearney and Wilson, 2018; Autor et al., 2019). If richer individuals choose to have fewer, "higher quality" children, and poorer individuals choose to have more, "lower quality" children, then this may translate into differences in college attendance. It is also possible that differences in health and income may affect the timing of childbirth, through differences in risky behavior. Finally, health and income differences may affect the likelihood of family formation, affecting the likelihood that the first-generation have children, or the household environment in which children are born (Clay et al., 2018).

[3]This estimate is comparable to the estimates provided by Isen et al. (2017) who examine the effects of the 1970 CAAA on first-generation later life earnings using the LEHD for 24 states.

We do not estimate any meaningful effects of prenatal pollution exposure on marital status. While being an outcome of interest in its own right, the absence of an effect on marital status also suggests the probability that a parent-child link is found — which depends somewhat on marital status — is unlikely to be correlated with exposure. We also estimate precise null effects of prenatal pollution exposure on the likelihood that first-generation individuals had any children, the number of children, or the timing of children. Collectively, these findings suggest that our second-generation estimates are unlikely to be driven by selection into the second-generation sample.

Finally, we explore whether parental investments in human capital could plausibly contribute to our results. We do this in two ways. First, we explore existing estimates of the contribution of cognitive and non-cognitive skills to college attendance. We calculate that a 10 $\mu g/m^3$ reduction in particulates would be equivalent to 0.2-0.25 Head Start Programs (Barr and Gibbs, 2017), a 2 standard deviation increase in self esteem, a 0.93 standard deviation reduction in impulsivity, a 0.73 standard deviation decrease in schooling problems, or a 0.38 standard deviation increase in cognitive ability (Lundberg, 2017). While each of these considerations alone is unlikely to drive our college attendance results it seems plausible that a combination of these considerations could be affected by parental investments as a residual explanation for our findings. To provide a more concrete sense as to whether parental investments in cognitive and non-cognitive skills is a plausible mechanism or not, we link the public use American Time Use Survey (ATUS) to administrative records on place of birth. This linkage allows us to integrate respondents from the ATUS into our existing data infrastructure. We provide evidence that treated parents — i.e., those that experienced better in-utero air quality — spend more time reading to their children. These findings suggest that parental investments in cognitive and non-cognitive skills contribute to the estimated college attendance increases. We caveat that the sample size is substantially smaller in this analysis.

Overall, our results suggest that the effects on college attendance reflect a combination

of parental resources and investments. Those directly affected by reduction in prenatal pollution exposure are richer and healthier. The parental earnings effects do not appear to be driven by educational attainment and so may reflect a broader health sequelae that may also have direct effects on the second-generation through increased "parenting human capital."

Our findings contribute to several literatures. First, we contribute to the 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). Specifically, we contribute to a nascent literature documenting that the economic consequences of early childhood and prenatal shocks persist across generational boundaries (Black et al., 2018; Barr and Gibbs, 2017; East et al., 2017; Akresh et al., 2018). We build on this literature by providing some insight into the mechanisms that could underlie the transmission of such shocks. Parental resources and investments, rather than through biological channels, drive our findings.

Second, we contribute to what is now a large literature on the Clean Air Act. While the question under study is novel, the quasi-experimental design described is well established. The 1970 CAAA variation (and similar variation from subsequent amendments) has been 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), and adult earnings (Isen et al., 2017). We follow the original research design, implemented by (Chay and Greenstone, 2003a) and (Isen et al., 2017), as closely as possible to discipline our analysis. Our results highlight that environmental regulations can have intergenerational consequences with important implications for human capital accumulation. Combining the first- and second-generation earnings effects delivers a total earnings effect of $8.6 billion for the 1972 cohort. The combined effect on cumulative lifetime earnings account for 76-95% of the monetized

damages associated with infant mortality, examined in previous work (Chay and Greenstone, 2003a).[4] To date the mortality effects of pollution have been the largest monetized damage in formal benefit-cost evaluations of environmental regulations. Mortality benefits currently account for 93% of EPA calculated benefits. As such, the absence of long-run and intergenerational considerations in benefit-cost analysis suggests that the returns to improvements in environmental quality may be substantially larger than previously thought. Most notably, we show that a within-generation approach to estimating the returns to improvements in environmental quality likely underestimates the total welfare effects.

Third, we contribute to the literature exploring economic mobility and opportunity (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). Within this literature the importance of neighborhoods has been established for the economic opportunities of children (Chetty et al., 2016; Chetty and Hendren, 2018a; Chyn, 2018; Deutscher, 2019). However, what it is about neighborhoods that matters is less clear. Can greater economic opportunity be created within neighborhoods, or should policy focus on encouraging households to move into neighborhoods with greater economic opportunity? Our results open up new lines of inquiry, raising questions about the potential role that investments in environmental quality may play in increasing upward mobility and economic opportunity.

The remainder of the paper is structured as follows. Section 2 presents background information and a conceptual framework, motivating the research design. Section 3 describes the data and empirical strategy. Section 4 presents our results. Section 5 discusses the implications of our results and concludes.

[4]The range of estimates arises because we bound the value of a statistical life between $8 and $10 million ($2016), following Lee and Taylor (2019)

2 Background and Conceptual Framework

What are the channels through which prenatal pollution exposure could have intergenerational effects on human capital? First, we have to understand how prenatal particulate matter exposure could directly affect health and development. In our empirical analysis we exploit variation in prenatal exposure to total suspended particles (TSP), the type of particulate matter that was regulated by the EPA at the time of the 1970s Clean Air Act amendments (CAAA). TSPs represent a complex mixture of organic and inorganic compounds found in the air that are smaller than 100 microns in diameter. TSPs enter the atmosphere due to both human activity, e.g., traffic, construction, and industrial production, and natural sources, e.g., dust and pollen.

All particulate matter can damage human health, although larger particles are less harmful than smaller particles. This is because larger particles settle more quickly than smaller particles and so are less likely to be inhaled. When larger particles are inhaled, they settle in the nose and throat and are usually expelled by coughing or sneezing. By contrast, smaller particles (those smaller than 10 microns) can remain in the air for extended periods of time, and once they are inhaled penetrate the lungs and the brain (through the olfactory bulb).

Once inside the body, particles are understood to affect respiratory function, lung development, and brain development. As particles can be transferred from the lungs and into the blood stream, they can also cause cardiovascular problems. These effects are amplified during gestation. Maternal exposure to particulate matter is understood to affect fetal health through a number of physiological pathways, such as oxidative stress, DNA damage to cells, damage to blood vessels, increased blood pressure, and reduced endothelial function (which could affect the transfer of nutrients to the fetus). Furthermore, particles can be transferred directly to the fetus through the bloodstream directly affecting respiratory, neurological and cardiovascular development (Dejmek et al., 1999; Glinianaia et al., 2004; Bové et al., 2019; Huifeng et al., 2020).

Particulate matter is understood to affect human capital and development throughout the

life cycle. It's effects are not limited to the gestational period, although marginal damages are understood to be larger during this period. This raises an important empirical challenge: we have to isolate the effects of prenatal pollution exposure from any contemporaneous effects experienced by the first-generation throughout their life cycle, or by the second-generation throughout their life cycle. Simply comparing the outcomes of individuals that experience prenatal exposure to locations with cleaner air to those of individuals who experience prenatal exposure to locations with dirtier air would not identify the effects of prenatal exposure because individuals born in "treatment" locations may be exposed to lower air over their entire life cycle, and so may their children. To distinguish between the effects of prenatal exposure and cumulative exposure within and across generations we require a research design that compares individuals who have different exposure to air pollution during gestation, but the same exposure post-birth. To formalize this idea we extend the framework presented by Isen et al. (2017).[5] In the process we highlight the mechanisms through which prenatal pollution exposure could (1) affect the economic outcomes of those directly affected and (2) transmit across generations to affect their children.

Let an individual's health stock be a function of inputs during two time periods: $h = h(p_1, p_2, X(p_1))$, where p_1 represents prenatal pollution exposure, p_2 represents post-birth pollution exposure, and X reflects their genetic endowment, which may be affected by prenatal pollution exposure through epigenetic effects — permanent changes in gene expression. Our research design exploits variation in the introduction of the 1970s clean air act amendments (CAAA), which lowered TSP concentrations in counties that exceeded the air quality standards, following its implementation. Moreover, we compare cohorts born just before and just after the CAAA came into effect (relative to cohorts born just before and after in counties that were unaffected). In this analysis the "treated" cohort born after the CAAA is exposed to lower p_1 and p_2. By contrast, cohorts born just before the CAAA had higher p_1 but the same exposure to p_2 (assuming they continue to live in their counties of birth).

[5]The foundations of this model are based on earlier work by Grossman (1972), Bleakley (2010), and Cunha and Heckman (2007).

By comparing these two groups, our analysis isolates the additional effect of changes in prenatal pollution exposure p_1. More precise details on the research design and econometric specification are presented in section 3.

2.1 First-Generation Effects

For the first-generation we model the effects of prenatal pollution on earnings, which is a function of the health stock h and education e,

$$y_f = y_f(e_f, h_f) = y_f(e_f(h_f(p_{f,1}, p_{f,2}, X(p_{f,1}))), h_f(p_{f,1}, p_{f,2}, X(p_{f,1})))$$

where y_f represents earnings and e_f represents years of schooling for the first-generation, those directly affected. The channels through which prenatal pollution exposure affects earnings can be characterized, by taking the total derivative of earnings, y_f, with respect to prenatal pollution exposure $p_{f,1}$,

$$\frac{dy_f}{dp_{f,1}} = \frac{\partial y_f}{\partial e_f}\frac{\partial e_f}{\partial h_f}\left[\frac{\partial h_f}{\partial p_{f,1}} + \frac{\partial h_f}{\partial X}\frac{\partial X}{\partial p_{f,1}}\right] + \frac{\partial y_f}{\partial h_f}\left[\frac{\partial h_f}{\partial p_{f,1}} + \frac{\partial h_f}{\partial X}\frac{\partial X}{\partial p_{f,1}}\right]$$

Prenatal pollution exposure, p_1, affects the health stock h, which in turn affects earnings directly ($\frac{\partial y}{\partial h}$) as well as indirectly through educational attainment ($\frac{\partial y}{\partial e}\frac{\partial e}{\partial h}$).

2.2 Second-Generation Effects

For the second-generation we model the effects of first-generation (parental) prenatal pollution exposure on second-generation educational attainment, which is a function of the second-generation's health stock h_s. The health stock of the second-generation is a function of second-generation prenatal pollution exposure, $p_{s,1}$, second-generation later-life pollution exposure, $p_{s,2}$, first-generation earnings, $y_f(h_f, e_f)$, first-generation education, $e_f(h_f)$, first-generation health stock, $h_f(p_{f,1}, p_{f,2}, X(p_{f,1}))$, and any genetic changes that may have occurred because of first-generation prenatal pollution exposure or second-generation prenatal

pollution exposure, $X(p_{f,1}, p_{s,1})$,

$$e_s(h_s(p_{s,1}, p_{s,2}, X(p_{f,1}, p_{s,1}); y_f(h_f(p_{f,1}, p_{f,2}, X(p_{f,1})), e_f(h_f(\cdot)); e_f(h_f(\cdot)); h_f(\cdot)))$$

The channels through which prenatal pollution exposure affects earnings can be characterized, by taking the total derivative of educational attainment, e_s, with respect to parental prenatal pollution exposure $p_{f,1}$,

$$\begin{aligned}
\frac{de_s}{dp_{f,1}} &= \frac{\partial e_s}{\partial h_s}\frac{\partial h_s}{\partial X}\frac{\partial X}{\partial p_{f,1}} && \text{(Biological Effects)} \\
&+ \frac{\partial e_s}{\partial y_f}\frac{\partial y_f}{\partial h_f}\left[\frac{\partial h_f}{\partial p_{f,1}} + \frac{\partial h_f}{\partial X}\frac{\partial X}{\partial p_{f,1}}\right] + \frac{\partial e_s}{\partial y_f}\frac{\partial y_f}{\partial e_f}\frac{\partial e_f}{\partial h_f}\left[\frac{\partial h_f}{\partial p_{f,1}} + \frac{\partial h_f}{\partial X}\frac{\partial X}{\partial p_{f,1}}\right] && \text{(Income Effects)} \\
&+ \frac{\partial e_s}{\partial e_f}\frac{\partial e_f}{\partial h_{f,1}}\left[\frac{\partial h_{f,1}}{\partial p_{f,1}} + \frac{\partial h_f}{\partial X}\frac{\partial X}{\partial p_{f,1}}\right] && \text{(Parental Education)} \\
&+ \frac{\partial e_s}{\partial h_f}\left[\frac{\partial h_f}{\partial p_{f,1}} + \frac{\partial h_f}{\partial X}\frac{\partial X}{\partial p_{f,1}}\right] && \text{(Parental Health)}
\end{aligned}$$

Broadly speaking, parental prenatal pollution exposure could affect the educational attainment of the second-generation, either through biological channels ($\frac{\partial e_s}{\partial h_s}$) or through the household environment, i.e, parental resources ($\frac{\partial e_s}{\partial y_f}$) and human capital ($\frac{\partial e_s}{\partial e_f}$ and $\frac{\partial e_s}{\partial h_f}$).

The remainder of this paper seeks to estimate $\frac{de_s}{dp_{f,1}}$, arising from changes in the level of parental prenatal particulate matter pollution exposure experienced at the time of the 1970 CAAA. We also analyze the mechanisms that help us to distinguish between the biological channels and changes in the household environment. Note that an implicit assumption underlying the decomposition of these mechanisms is that $p_{f,2}$, $p_{s,1}$, and $p_{s,2}$ are not a function of $p_{f,1}$. If, for example, parental prenatal pollution exposure is positively correlated with pollution exposure in later life then we can't identify the effects of parental prenatal pollution exposure separately from later-life first-generation exposure or second-generation exposure. The following section provides more detail about the research design and empirical

specifications that help us to isolate that the effects of parental prenatal pollution exposure, as well as providing direct evidence that there is no differential effect of prenatal pollution exposure on later life pollution exposure, i.e, $\frac{\partial p_{f,2}}{\partial p_{f,1}} = 0$, $\frac{\partial p_{s,1}}{\partial p_{f,1}} = 0$, and $\frac{\partial p_{s,2}}{\partial p_{f,1}} = 0$.

3 Research Design

In this section we provide an overview of the data and sample construction process, as well as the empirical specifications used to estimate the effects of parental prenatal particulate matter exposure on second-generation outcomes.

3.1 Data

To study the intergenerational effects of pollution exposure, it is necessary to identify the location and date of birth for each parent, infer their exposure to ambient air pollution, link these parents to their children, and measure outcomes. No single dataset has all of these features, and so our analysis requires linking decennial Census, administrative records, and survey data. This linkage is done using unique anonymous personal identifiers called Protected Identification Keys (PIKs), which can be thought of as "scrambled" Social Security Numbers. PIKs 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.[6]

3.1.1 Parent-Child Links

We begin by assembling a database of 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 decennial

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

Census from 2000 and 2010, and the American Community Survey (ACS) from 2005–2015.[7] 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. We focus on two sets of these links — "certain" links, where the survey and Census information allows us to identify a parent-child link with certainty — and "probable" links, where the survey and Census information allows us to identify the spouse of a certain link parent, who are likely also parents (but may be step-parents).

Each dataset contains slightly different information on relationships within households. The decennial Census and ACS data both contain detailed information on relationships within the household, with one important limitation — the Decennial Census/ACS relationship question asks for information only on the relationship between an individual and the head of household. 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. In the main analysis we focus on certain parent-child links to minimize any attenuation bias introduced through the incorrect assignment of children to affected, or unaffected, parents. In the appendix materials we show that our results are robust to incorporating probabilistic links.

3.1.2 Pollution Exposure

To analyze the intergenerational effects of pollution exposure, we need to infer the level of ambient air pollution that parents were exposed to during gestation. We do this in three

[7]Other probable links could in principle be measured in the universe of IRS 1040 tax returns, however we focus on the Census and ACS links as we can infer with certainty these links are between parents and children.

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 we aggregate to the county level, and link these county-level measures to the parents' place of birth.

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. The Numident contains information on individuals' exact date of birth, and place of birth. As place of birth information is not standardized, we assign county of birth information to individuals using a crosswalk, provided by Isen et al. (2017) combined with a probabilistic matching approach used in Voorheis (2017). Using this approach we identify county of birth for first and second generation individuals. For cases where county of birth cannot be assigned using the Isen et al. (2017), we attempt to match the Numident place of birth string to the United States Geological Survey's list of places using a fuzzy string matching algorithm based on the optimal string alignment (OSA) distance metric.

With information about the place of parents' birth in hand, we infer the level of pollution exposure experienced by these individuals based on average pollution concentrations during the 270 days of gestation 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[8]. Our pollutant of interest is particulate matter. For the relevant period of time (around 1970), the primary regulatory definition of particulates was total suspended particles (TSP), defined as particulate matter with a density of less than 50 microns, measured in units $\mu g/m^3$. We retrieve data from all TSP monitors between 1960–1980.

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

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

TSP in that county on that day. We then calculate county-level gestational period averages of prenatal TSP exposure for each birth between 1960 and 1980.

3.1.3 Outcomes

Finally, we require information on outcomes for parents and children, as well as other information on observable 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 sample of households (about 1 percent per year), we observe outcomes from only a fraction of the parents and children identified above.

Our main outcome of interest is college attendance for the second generation. We focus on this outcome because we have limited data on earnings, given the timing of when the first and second generations are born. In the absence of earnings data we believe that college attendance data is the most important and interesting outcome to explore. We also explore other measure of human capital attainment, such as high school completion. The reason we focus on human capital accumulation is because 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. Human capital measures available in the ACS are prevalent among people under the age of 22, who are much more likely to have parents that were affected by the 1970 CAAA. Additionally, although some second generation individuals would have earnings in IRS administrative records, there are very few, if any, individuals around the "overtaking age" when earnings are near the lifecycle peak.[9]

We also explore several outcomes for the first generation with a view to understanding how the first-generation is affected and the mechanisms through which the consequences of prenatal pollution exposure could be transmitted across generations. We explore family

[9]For instance, in 2018, the only 30 year olds with parents born after the CAAA would have been born when their parents were 16.

structure using information on marital status and fertility. We are interested in exploring these outcomes for two reasons. First, we believe that family structure could plausibly affect parental investments and resources available to children, which could be an important mechanism for intergenerational transmission. Second, we are interested in the potential for selection into the second generation sample — pollution exposure could affect the likelihood, or timing, of children. The ACS microdata contains information on the 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 also construct variables providing information on the timing of children. Beyond changes in family structure we also explore economic outcomes, defining variables for unemployment, public assistance receipts, and wages from detailed ACS questions on income and labor force participation.

The ACS also provides sociodemographic information for the second generation, including race, sex, and age. The first and second generation do not always appear in the ACS at the same time and so we also collect demographic characteristics from the decennial Census to control for parental characteristics when evaluating second-generation outcomes. We also collect information on the characteristics of first- and second-generation county of birth — population, employment, personal income per capita and total transfer income — from the Bureau of Economic Analysis' Regional Product Accounts.

3.2 Empirical Strategy

We are interested in estimating the relationship between prenatal particulate matter exposure and the later life outcomes of the children of affected individuals 40-50 years later — the second generation. Our baseline model takes the following form,

$$Outcome_{i,j,c,t} = \beta_0 + \beta_1 TSP_{j,c,t} + \gamma X_j' + \delta X_c' t + \alpha_c^j + \alpha_{st}^j + \epsilon_{i,j,c,t,y} \quad (1)$$

Outcomes are measured for child i born to parent j. Exposure is measured for parent

j. Specifically, $TSP_{j,c,t}$ 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. $\mathrm{X}_c t$ is a vector of county-level characteristics, measured in 1969, interacted with linear and quadratic time trends. α_c^j are county-of-birth fixed effects that control for time-invariant unobserved determinants of the labor market outcomes and family structure for parents born in county c. α_{st}^j are birth-state $\times$ birth-year fixed effects which control for time-varying determinants of the long-run outcomes, common across all parents born in a state s in year t. The coefficient of interest is β_1 which reflects the effect of a one-unit increase in parental, prenatal, TSP exposure on child j's later life outcomes. In extensions we include second-generation controls, including the second-generation birth-county fixed effects, α_c^i, second generation birth-state $\times$ birth-year fixed effects, α_{st}^i, and second-generation individual characteristics, X_i'. Across all specifications we cluster our standard errors by the first generation's county of birth — the level at which we measure exposure.

There are two key threats to identification that need to be addressed. First, it is highly likely that exposure to particulate matter is correlated with many observable and unobservable characteristics that are correlated with long-run economic and social outcomes. This is a standard identification when trying to identify the effects of endogenous pollution on any 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. For this reason, we use an instrumental variables design.

The second identification issue is specific to identifying the effects of early-life shocks on later life outcomes. Even if the shock is as good as randomly assigned one would want to show that prenatal exposure to a shock does not affect later life exposure. This was the focus of our discussion in the conceptual framework. If differences in early-life pollution exposure

(or any early life shock) are also correlated with differences in later life pollution, it is not possible to disentangle the persistent effect of early-life shocks from later life exposures. We argue that our choice of instrumental variable allows us to address this issue.

3.2.1 Using the 1970 CAAA in an Instrumental Variables Design

To address the endogeneity concerns related to pollution exposure and isolate the effects of early-life pollution exposure, we instrument for changes in particulate matter exposure using the introduction of the 1970 Clean Air Act Amendments. 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.

The NAAQS standards were applied at the level of Air Quality Control Regions (AQCRs), which were EPA-defined collections of counties. AQCRs that exceeded these standards were considered to be nonattainment areas. As a consequence of being designated a nonattainment area, a plan must be implemented to meet the standard. The EPA defined an AQCR as being in nonattainment of NAAQS if any of the individual monitors in the AQCR fell outside the NAAQS standards. Thus effectively the NAAQS standards bind at the county level. Counties with monitors out of attainment needed to make large improvements in air quality, while counties which were in attainment did not, even if they were located in an AQCR which was out of attainment (in other words, attainment counties in nonattainment AQCRs were inframarginal). Thus, the relevant fact for us is whether a county was in nonattainment

of the NAAQS standards; however, the EPA did not publish county-level nonattainment designations until the late 1970s.[10] Thus we follow the existing literature in modelling the effect of the Clean Air Act using the nonattainment status of counties, not AQCRs, noting that it is the bindingness of EPA regulations, and not the nonattainment designations themselves, which produce our quasi-experimental variation in pollution exposure.

Following the existing literature we reconstruct which counties would have been in nonattainment using TSP concentrations from the years before the CAAA was enacted.[11] Nonattainment of the primary air quality standard for TSP set in the 1970 CAAA occur 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 one monitor in 1970. This allows us to categorize 258 counties as "nonattainment" counties, and 319 counties as "attainment" counties.[12]

We model the change in air pollution using an indicator variable for county nonattainment 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-differences regression model,

$$TSP_{j,c,t} = \alpha_0 + \alpha_1(Nonattainment_{c,1970} \times 1[\tau > 1971]) + \alpha_c^j + \alpha_{st}^j + \gamma X_j' + \delta X_c' t + \nu_{j,c,t} \quad (2)$$

where TSP exposure for parent j in county c in year t is regressed on a time-invariant county indicator equal to 1 if a county is designated as nonattainment, $Nonattainment_{c,1970}$,

[10] Note also that though in principle it would possible to define an AQCR-level treatment, essentially all monitored AQCRs had at least one monitor out of attainment, rendering this approach futile in practice.

[11] The EPA did not make county-level nonattainment designations publicly available as noted above. However, TSP air quality standards are known, and so we can reconstruct which counties would have been in nonattainment using monitor-level data

[12] 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.

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 nonattainment counties following the implementation of the 1970 CAAA. The parameter of interest is α_1, which provides a difference-in-differences estimate of the impact of nonattainment designation on prenatal TSP exposure in the years after CAAA regulations went into place. This research design also isolates the effects of prenatal pollution exposure. The introduction of the 1970s CAAA means that we are comparing cohorts who have different exposure to TSPs during gestation, but the same exposure post-birth. Table 2 shows while the 1970s CAAA reduced first-generation prenatal exposure, there is no statistically significant effect on first-generation later life exposure (column 2), second-generation prenatal exposure (column 3), or second-generation later-life exposure (column 4). In Table A2 we also show that there is little evidence of differential migration. We estimate that exposure to the 1970 CAAA is associated with small reductions in the likelihood that the first-generation ever migrates out of their county of birth, corresponding to a 0.78 percentage point reduction. However, we do not estimate any effect on the likelihood that the first-generation moves out of state. We also do not see any evidence that the second-generation migrate away from their parents county-of-birth, indicating that first-generation movers are those without children. This suggests that differential migration is unlikely to be an important concern. Ultimately, this is reflected in the fact that there is no differential exposure to pollution throughout the life-cycle.

In the second stage, we use the predicted TSP levels from equation 2 in place of observed TSP levels,

$$Outcome_{i,j,c,t} = \rho_0 + \rho_1 \widehat{TSP}_{j,c,t} + \alpha_c^j + \alpha_{st}^j + \gamma X_i' + \delta X_c' t + \varepsilon_{i,j,c,t,y} \tag{3}$$

where the coefficient of interest ρ_1 captures the effect of a one-unit increase in CAAA-driven parental, prenatal, TSP exposure on child j's later life outcomes.

We show, consistent with previous research on the Clean Air Act, that the first stage relationship is strong — that nonattainment status is associated with significant and persistent

declines in particulate matter concentrations in the years after the 1970 CAAA came into effect. Table 1 presents the first-stage relationship for our college attendance sample. We observe across all specifications that nonattainment is associated with an 8 $\mu g/m^3$ reduction in TSP, on average. In Figure 1 we explore the effects of nonattainment on prenatal TSP exposure in an event study framework, where nonattainment has separate effects in each year. Prior to the introduction of the 1970s CAAA we find no statistically significant, or economically meaningful, differences between nonattainment and attainment counties, providing support for the parallel trends assumption. Following implementation, we estimate an immediate and persistent reduction in prenatal TSP exposure, as would be expected.

In addition to providing evidence in support of the parallel trends assumption we show that there are limited differences between attainment and nonattainment county characteristics prior to the 1970 CAAA (Table A1). Isen et al. (2017) also show that a nonattainment designation is not correlated with any changes to the observable characteristics of mothers that gave birth in the years following the 1970 CAAA.

The main concern regarding the instrumental variable research design is the exclusion restriction. To identify the intergenerational effects of TSP exposure it must be the case that the CAAA doesn't affect outcomes in any other way than through reductions in pollution. Isen et al. (2017) make the point that nonattainment designations could affect economic competitiveness (Greenstone, 2002; Greenstone et al., 2012; Walker, 2011, 2013). However, existing evidence suggests that the effects on the broader local economy are small, affecting less than 0.7 percent of the total workforce and is based on the 1990 Clean Air Act Amendments (Walker, 2013). As the 1970 CAAA was the first major regulation to be introduced, actions to reduce emissions may have been less costly in 1970 due to low-hanging fruit, than in the 1990s, attenuating concerns about economic competitiveness. Nevertheless, it is possible that the 1970 CAAA contributed to a decline in economic conditions for nonattainment counties, affecting the long-run economic prospects of affected individuals. As competitiveness effects would be expected to have the opposite effects on health to reduc-

tions in pollution exposure, our 2SLS estimates will understate the intergenerational effects of pollution exposure if the exclusion restriction is violated. The reduced form effect of nonattainment remains valid and is interpreted as the intergenerational effects of the 1970 Clean Air Act Amendments, rather than pollution. Our reduced form and corresponding 2SLS estimates produce conceptually identical results, although the signs of the effects are different – nonattainment increases second generation human capital, and pollution exposure decreases second generation human capital.

4 Results

Table 3 presents the results of our analysis on the effects of parental, prenatal, pollution exposure on the likelihood that their child attends college, using our preferred second-generation sample that uses certain parent-child links. We estimate that a 10 $\mu g/m^3$ reduction in first generation prenatal pollution exposure is associated with a 3.2-3.8 percentage point increase in the likelihood that the second-generation attends college.[13] Relative to the mean, this corresponds to an 8% increase in college attendance. This is a substantial effect, equivalent in magnitude to the second-generation effect of 0.2-0.25 Head Start Programs (Barr and Gibbs, 2017). In Panel B of Table 3 we present reduced form effects, capturing the effect of the 1970 CAAA on second-generation outcomes. We estimate that the reduced form effect of the 1970 CAAA is associated with a 2.6-3 percentage point increase in the likelihood of college attendance. Using the average nonattainment induced changes in TSP for this sample ($\approx 8\mu g/m^3$) our IV estimates imply a 2.6-3.08 percentage point increase in the likelihood of attending college. The 2SLS and reduced form estimates are very similar.

We also explore whether parental prenatal TSP exposure affects the likelihood of dropping out of high school. In Table A3 we estimate that a 10 $\mu g/m^3$ reduction in first generation prenatal pollution exposure is associated with a 0.5 percentage point reduction in the likeli-

[13] In Tables A7, A7, and A9 we estimate the effects defining exposure based on whether it is maternal, paternal, or if both parents were exposed. We estimate larger effects if both parents were exposed, but the combined effect is smaller than the sum of mother-only or father-only effect.

hood that their child drops out of high school. While the absolute effect is smaller, dropping out of high school is much rarer than attending college (6% of our sample drop out of high school). A 0.5 percentage point reduction in the likelihood of dropping out of high school is also an 8% reduction, relative to the mean.

While we are not able to examine the intergenerational effects of prenatal pollution exposure on earnings, the effects on college attendance plausibly have meaningful impacts on the earnings potential of second-generation individuals. Using a college wage premium of $28,000 p.a. (estimated using the public use CPS-ASEC), combined with a graduation rate of 50 percent to convert college attendance to college completion[14], and a 3 percent real discount rate (5 percent discount rate + 2 percent wage growth) we predict a $\$1{,}515/10\mu g/m^3$ increase in second generation cumulative lifetime earnings at age zero of the first generation. Combining the number of children born in nonattainment counties in 1972 (1.52 million), the probability that one of these individuals had a child (0.63), and the average number of children that were born to these individual (2.24), we calculate that the total second generation earnings effects for the 1972 cohort is around $3.2 billion dollars. If reductions in TSP were persistent (and evidence suggests they were) we'd expect these benefits to be realized for each cohort since 1972. If the intergenerational effects of prenatal pollution exposure have effects on income other than through college this number represents a lower-bound estimate of the second-generation benefits.

Our results are robust to including probabilistic links, although the estimates are marginally smaller, consistent with the introduction of classical measurement error (Table A6). Our results are also robust to reducing the time-window over which we estimate the effects (Tables A4 and A5). Using an alternative RDD research design (discussed in Appendix A.5) we estimate that nonattainment is associated with a 2-4 percentage point increase in the likelihood of second-generation college attendance (Table A10).

[14]See https://nces.ed.gov/programs/coe/indicator_ctr.asp

4.1 Exploring Mechanisms

What are the mechanisms through which prenatal pollution exposure propagates across generations. We posit that there are two broad channels through which these effects could arise. The first channel is biological. Gestational pollution exposure could result in epigenetic changes, i.e., permanent changes in gene expression. If so, hereditary transmission of these changes could affect the second generation. The second mechanism is economic. Gestational pollution exposure affects the human capital and earnings of the first generation. In turn, parental resources and investments may affect the likelihood of college attendance for their children.

4.1.1 Biological Effects

To explore the empirical relevance of a hereditary mechanism we examine the differential effect of parental, prenatal, pollution exposure on adopted and biological children. This exercise allows us to examine the degree to which permanent changes in gene expression, triggered by prenatal pollution exposure, are passed down from parent to child, affecting human capital, and in turn the likelihood of attending, or 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 hereditary mechanism.

Biological children could be affected through both changes in household environment (ν) and through genetic channels (γ). By contrast, adopted children are only be affected through the household environment (ν) — i.e. parental resources and investments. 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 make differential investments 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.

There are two potential concerns. First, one may be concerned that there is a differential propensity to adopt or not adopt children in nonattainment counties, resulting in selection into the second-generation adopted sample. In column 1 of Table 4 we estimate that there is no statistically significant or quantitatively meaningful effect of prenatal TSP exposure on the likelihood of adopting a child. Second, adopted children may be affected through hereditary channels if their biological parents were born in nonattainment 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 argue is unlikely to be the case in a systematic way. In support of this assumption, we observe that adopted children are more likely to be born in a different county from their adopted parents than biological children.

The remaining results in Table 4 explore whether there is a differential effect of prenatal pollution exposure on second-generation college attendance for adopted versus biological children. Across all specification we estimate no statistically significant, differential effects of parental pollution exposure. As a robustness check we explore differences between step-children and biological children. We do not estimate any statistically significant differential effects of parental pollution exposure between these groups (Table B1). This suggests that the estimated college attendance effect likely arises due to parental resources and investments, rather than genetic pathways (Bjorklund and Chadwick, 2003; Bjorklund, 2006; Bjorklund et al., 2007, 2010).

4.1.2 Income Effects

Having found little evidence to support a biological mechanism we turn to economic channels. First, we explore the effects of prenatal pollution exposure on labor market outcomes. Table 5 presents the results of this analysis.

We estimate that a 10 $\mu g/m^3$ reduction in in-utero exposure to TSP is associated with a 1.2 percent increase in earnings at the time of ACS response. Evaluated at the mean earnings, this corresponds to an annual earnings effect of $367 compared to the control mean. This effect is similar in magnitude to the estimates presented in Isen et al. (2017). This is encouraging for two reasons. First, we are using the same methodology. Second, our findings highlight the external validity of the Isen et al. (2017) findings, that are restricted to 24 states. Our sample contains individuals in all states which had active particulate monitors in 1970, and does not restrict the age of respondents to 29-31.

Assuming that the estimated earnings effect is constant over the life cycle, and that earnings are discounted at a real rate of 3 percent back to age zero, we calculate that the lifetime earnings effect of a 10 $\mu g/m^3$ increase in TSP is $3,499 per person ($5.3 billion for the 1972 cohort).

We also explore effects on labor force participation (column 2), unemployment (column 3), and public assistance (column 4). We fail to reject the null hypothesis that there is no relationship between gestational TSP exposure and these outcomes. The absence of any effects suggest that TSP exposure does not appear to affect labor market outcomes on the extensive margin. This is consistent with Isen et al. (2017) who estimate increases in the number of quarters worked, rather than extensive margin labor force participation effects. Finally, in Table B4 we explore first-generation educational outcomes. We fail to reject the null hypothesis that prenatal TSP exposure has any effect on years of schooling or college attendance. This is consistent with Isen et al. (2017) who argue that the effects on earnings are too small to be driven by schooling. Collectively, these results suggest that the earnings effects likely reflect a broader health sequelae.

The estimated income effects imply a relationship between parental income and college attendance, 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. If our findings were driven entirely by the increase in parental earnings then a $10,000 increase in household wealth would be associated with an 9.2-11 percentage point increase in the 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. An additional limitation of the existing literature is that wealth shocks are experienced much later in life. Potential college attendees are usually in their teenage years at the time that the wealth shock is realized. For example, Lovenheim (2011) and Lovenheim and Lockwood Reynolds (2013) look at wealth shocks in the 4 years prior to turning 18, and most of the oldest child in Bulman et al. (2017) are teenagers at the time that their parents win the lottery. It is entirely possible that wealth shocks during the teenage years have much less effect on college attendance. Increased earnings earlier in the life cycle may be complementary with human capital investments affecting the educational trajectory that children take. In our sample the second-generation experience the wealth shock from birth and so it is possible that this translates into a more elastic college-earnings relationship.

4.1.3 Family Structure and Fertility Effects

In light of the income effects it is possible that part of the college attendance effect reflects selection into the second-generation sample through fertility effects. If treated individuals are richer and healthier this may affect the likelihood of family formation, through marriage and fertility choices. Becker (1960) posits that wealthier parents may have a preference for quality over quantity of children. However, empirical evidence also suggests that increases in earnings or wealth are associated with increases in fertility (Lovenheim and Mumford, 2013; Black et al., 2013; Dettling and Kearney, 2014; Kearney and Wilson, 2018; Autor et al., 2019). As such, we might expect parents that were exposed to lower levels of TSP to have smaller families. If family size is negatively correlated with the likelihood of college attendance then this could explain part of our results. It is also possible that pollution exposure could affect fertility or fecundity (Carre et al., 2017). In such a case there may be reduced selection into the second-generation sample for the control group.

Table 6 presents our findings, exploring the family structure effects of prenatal exposure to TSP. We estimate that there are few meaningful changes in family structure. We fail to reject the null hypothesis that there is any change in the likelihood of being married. We also fail to reject the null hypothesis that prenatal pollution exposure has no effect on the likelihood of having any children, the number of children, or the timing of children.

One might be concerned that the average effect may be a net zero rather than a true zero if income effects have a differential fertility response for men and women. Economic theory suggests that improvements in male labor market conditions should be associated with increases in fertility, while better wages and employment opportunities for women should have opposing income and substitution effects (Schaller, 2016). In Tables B2 and B3 we explore wehther there are differential family structure effects by sex. We do not find any effect of particulate matter reductions on the likelihood that women get married, on the number of children, or the age at first birth. The effects are small in magnitude. For men, we estimate that a 10 $\mu g/m^3$ reduction in prenatal TSP exposure is associated with a 0.6

percentage point decrease in the likelihood of getting married and 0.025 fewer children (a 2.3% reduction relative to the mean). Both effects are statistically significant at the 10% level, economically small, and contrast with what economic theory would predict. As such, we do not see these changes as having economically meaningful effects on fertility or family structure.

In Table B5 we explore whether reductions in particulate matter exposure had effects on who individuals married, an examination of assortative matching. We estimate that a 10 $\mu g/m^3$ reduction in prenatal TSP exposure is associated with a 64 percentage point increase in the likelihood that an individuals' partner was treated, but no meaningful effects on the characteristics of partners. These findings are interesting because it suggests that our analysis of individual income effects likely understates the increase in household earnings. If both parents were affected a \$10,000 increase in household wealth would instead predict a 4.6-9.5% increase in the likelihood of attending college.[15] This is still a large effect, however, in light of the caveats discussed above (i.e., general equilibrium effects + earlier life exposure to household wealth increase) it is possible that parental resources are an important driver of the college effects.

Collectively, these findings suggest that our results are unlikely to be driven by differential selection into the second generation sample and give further support to the role that parental resources might play in driving the college attendance effects..

4.1.4 Parental Investments

Traditionally, models exploring the mechanisms underlying intergenerational persistence have implicitly assumed that all parents as equally good. More recently, Cunha and Heckman (2007), Heckman and Mosso (2014), and Becker et al. (2018) model increases in parental human capital as having an effect not only on the earnings of parents but also the production of their children's human capital, introducing a new mechanism — parental investments.

[15] We bound the effects using the smallest college effect in Table 3 and the largest effect in Table A9.

If reductions in prenatal pollution exposure increased 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, and in turn improve child outcomes (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). With the data available it is difficult to directly evaluate this consideration. 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. Barr and Gibbs (2017) explore the intergenerational consequences of the Head Start program and estimate that it increased the likelihood of second-generation college attendance by 15 percentage points. As such, a $10\mu g/m^3$ reduction in TSP has an equivalent effect on college attendance to 0.2-0.25 Head Start Programs. Lundberg (2017) explores how specific non-cognitive and cognitive skills are associated with college attendance. We calculate that a $10\mu g/m^3$ reduction in TSP would be equivalent to a 2 standard deviation increase in self esteem, a 0.93 standard deviation reduction in impulsivity, a 0.73 standard deviation decrease in schooling problems and a 0.38 standard deviation increase in cognitive ability. These are meaningful effects. As such it is possible that parental investments in cognitive and non-cognitive skills could contribute at least part of our results.

One concrete approach to understanding the relevance that parental investments in cognitive and non-cognitive skills may play is to examine the relationship between prenatal pollution exposure and parental time-use. We do this by linking the public use American Time-Use Survey with our existing data infrastructure. We caveat that sample size is substantially smaller in this analysis because we are restricted to ATUS respondents born in monitor-counties during our evaluation window.

We explore the effects of prenatal pollution exposure on the time spent on educational activities with their children. We look at the effects of parental exposure on the time spent reading with children, the time spent on educational activities, and the total time spent with

children. We find that reductions in prenatal TSP exposure are associated with an increase in time spent reading to children. Our estimate corresponds to an average increase of 1.5 minutes per day for each 10 $\mu g/m^3$ decrease in TSP. This is a 54% increase relative to the control mean. Effects on time spent on other educational activities and total time spent with children point in the same direction, however, they are less precisely estimated. Taken at face value, the central estimates suggest a 17% increase in time spent on educational activities and a 4.5% increase in the total time spent with children.

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, 2014; Massaro, 2017; Mendelsohn et al., 2018). We do not claim that the college effects are necessarily driven by reading alone. Instead, it is likely that parents who spent time reading to their children are also closely engaged with them along multiple dimensions. Consequently, we argue that the estimated college attendance effects are likely to arise, at least in part, because of parental investments and choices that affect cognitive and non-cognitive skills.

5 Conclusion

We provide early quasi-experimental evidence on the intergenerational consequences of prenatal 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 the children of those that were were directly affected experienced substantial increases in the likelihood of attending college 40-50 years later. We present evidence to suggest that the intergenerational transmission of this early-life shock to parental endowments arises through increased parental resources and investments.

Back-of-the-envelope calculations suggest that the combined cumulative lifetime earnings effects for the 1972 cohort are comparable in magnitude to the mortality benefits estimated

in Chay and Greenstone (2003a). To date, the monetized mortality benefits of environmental regulations constitute the vast majority of overall benefits in benefit-cost analysis. We argue that accounting for intergenerational effects would substantially increase the returns to investments in environmental quality. Within-generation estimates of marginal damages substantially underestimate the total welfare effects of improving environmental quality.

It is striking that these effects arise from such short exposure times, early in life. If relatively low levels of air pollution have intergenerational consequences through parental resources and investments, it is likely that larger shocks and policies also have intergenerational consequences. This reinforces the potential welfare benefits that could be realized through a reallocation of resources from later to earlier in the life-cycle. Understanding the margins through which such reallocations can be delivered remains an important area for research. Our findings suggest that investments in environmental quality could be an important contribution to such efforts.

In turn, our results have important implications for inequality and economic opportunity. It is well established that economic and environmental inequality walk hand-in-hand. Individuals who live in low-income areas and disadvantaged communities are exposed to higher levels of pollution than those that live in high-income areas (Commission for Racial Justice, United Church of Christ, 1987; Mohai et al., 2009; Banzhaf et al., 2019; Colmer et al., 2020; Currie et al., 2020). Early-life exposure to pollution has intergenerational consequences that are propagated through economic resources and investments. As such our results open up new lines of inquiry, suggesting that environmental quality may have important implications for upward mobility and economic opportunity.

References

Akee, R., W. Copeland, E.J. Costello, and E. Simeonova, "How Does Household Income Affect Child Personality Traits and Behaviors?," *American Economic Review*, 2018, *108* (3), 775–827.

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

25265, 2018.

Almond, D., "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*, 2006.

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

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

Autor, D., D. Dorn, and G. Hanson, "When Work Disappears: Manufacturing Decline and the Falling Marriage Market Value of Young Men," *American Economic Review: Insights*, 2019, *1* (2), 161–178.

Bailey, M., H. Hoynes, M. Rossin-Slater, and R. Walker, "Is the Social Safety Net a Long-Term Investment? Large-Scale Evidence from the Food Stamps Program," *NBER Working Paper 26942*, 2020.

Banzhaf, S., L. Ma, and C. Timmins, "Environmental justice: The economics of race, place, and pollution," *Journal of Economic Perspectives*, 2019, *33* (1), 185–208.

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

Bastian, J. and K. Michelmore, "The Long-Term Impact of the Earned Income Tax Credit on Children's Education and Employment Outcomes," *Journal of Labor Economics*, 2018, *36* (4), 1127–1163.

Becker, G., *Demographic and Economic Change in Developed Countries.*, Princeton: Princeton University Press, 1960.

_ and N. Tomes, "An Equilibrium Theory of the Distribution of Income and Intergenerational Mobility," *Journal of Political Economy*, 1979.

_ , S.D. Kominers, K.M. Murphy, and J.L. Spenkuch, "A Theory of Intergenerational Mobility," *Journal of Political Economy*, 2018, *126* (S1), S7–S25.

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

Bjorklund, A., "The Origins of Intergenerational Associations: Lessons from Swedish Adoption Data," *Quarterly Journal of Economics*, 2006, *121* (3), 999–1028.

_ and L. Chadwick, "Intergenerational Income Mobility in Permanent and Separated Families," *Economic Letters*, 2003, *80* (2), 239–246.

_ , **L. Lindahl, and M. Lindquist**, "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*, 2010, *10* (1).

_ , **M. Jantti, and G. Solon**, "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*, 2007, *7* (2).

Black, D., N. Kolesnikova, S.G. Sanders, and L.J. Taylor, "Are Children "Normal"?," *Review of Economics and Statistics*, 2013, *95* (1), 21–33.

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

_ **and P. Devereux**, "Recent Developments in Intergenerational Mobility," in "Handbook of Labor Economics," Elsevier, 2011.

_ , _ , **and K. Salvanes**, "Why the Apple doesn't fall far: Understanding Intergenerational Transmission of Human Capital," *American Economic Review*, 2005.

_ , _ , **and** _ , "From the Cradle to the Labor Market? The Effect of Birth Weight on Adult Outcomes," *Quarterly Journal of Economics*, 2007.

Bleakley, H., "Health, Human Capital, and Development," *Annual Review of Economics*, 2010, *2*, 283–310.

Bové, H., E. Bongaerts, and E. et al. Slenders, "Ambient black carbon particles reach the fetal side of human placenta," *Nature Communications*, 2019, *10* (3866).

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

Carre, J., N. Gatimel, J. Moreau, J. Parinaud, and R. Leandri, "Does air pollution play a role in infertility? A systematic review," *Environmental Health*, 2017, *16* (82).

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

_ **and** _ , "The Impact of Air Pollution on Infant Mortality: Evidence from Geographic Variation in Pollution Shocks Induced by a Recession," *The Quarterly Journal of Economics*, 2003, *118* (3), 1121–1167.

_ **and** _ , "Does Air Quality matter? Evidence from the Housing Market," *Journal of Political Economy*, 2005.

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

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

_ **and** _ , "The Effects of Neighborhoods on Intergenerational Mobility II: County Level Estimates," *Quarterly Journal of Economics*, 2018.

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

_ **, N. Hendren, and L. Katz**, "The Effects of Exposure to Better Neighborhoods on Children: New Evidence from the Moving to Opportunity Experiment," *American Economic Review*, 2016, *106* (4), 855–902.

_ **,** _ **, P. Kline, and E. Saez**, "Where is the land of Opportunity? The Geography of Intergenerational Mobility in the United States *," *The Quarterly Journal of Economics*, 2014, *129* (4), 1553–1623.

Chyn, E., "Moved to Opportunity: The Long-Run Effect of Public Housing Demolition on the Labor Market Outcomes of Children," *American Economic Review*, 2018, *108* (10), 3028–56.

Clay, K., M. Portnykh, and E. Severnini, "Toxic Truth: Lead and Fertility," *NBER Working Paper 24607*, 2018.

Colmer, J., I. Hardman, J. Shimshack, and J. Voorheis, "Disparities in PM2.5 air pollution in the United States," *Science*, 2020, *369* (6503), 575–578.

Commission for Racial Justice, United Church of Christ, "Toxic Wastes and Race in the United States: A National Report on the Racial and Socio-Economic Characteristics of Communities with Hazardous Waste Sites," 1987.

Cunha, F. and J. Heckman, "The Technology of Skill Formation," *American Economic Review*, 2007, *97* (2), 31–47.

_ **,** _ **, and S. Schennach**, "Estimating the Technology of Cognitive and Noncognitive Skill Formation," *Econometrica*, 2010, *78* (3), 883–931.

Currie, J., J. Voorheis, and R. Walker, "What Caused Racial Disparities in Particulate Exposure to Fall? New Evidence from the Clean Air Act and Satellite-Based Measures of Air Quality," *NBER Working Paper 26659*, 2020.

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

Dejmek, S., G. Selevan, I. Benes, I. Solansky, and R.J. Sram, "Fetal growth and maternal exposure to particulate matter during pregnancy," *Environmental Health Perspectives*, 1999, *107* (6), 475–480.

Dettling, L.J. and M. Kearney, "House Prices and Birth Rates: The Impact of the Real Estate Market on the Decision to have a Baby," *Journal of Public Economics*, 2014, *110*, 82–100.

Deutscher, N., "Place, Peers, and the Teenage Years: Long-Run Neighborhood Effects in Australia," *American Economic Journal: Applied Economics*, 2019.

East, C., S. Miller, M. Page, and L. Wherry, "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*, 2017.

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

Glinianaia, S., J. Ranking, J. Bell, T. Pless-Mulloli, and D. Howel, "Particulate Air Pollution and Fetal Health: A Systematic Review of the Epidemiologic Evidence," *Epidemiology*, 2004, *15* (1), 36–45.

Greenstone, M., "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*, 2002.

_ **, J. List, and C. Syverson**, "The Effects of Environmental Regulation on the Competitiveness of US Manufacturing," *Mimeo*, 2012.

Grossman, M., "On the Concept of Health Capital and the Demand for Health," *Journal of Political Economy*, 1972, *80* (2), 223–255.

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

_ **and S. Mosso**, "The Economics of Human Development and Social Mobility," *Annual Review of Economics*, 2014, *6* (1), 689–733.

_ **, R. Pinto, and P. Savelyev**, "Understanding the Mechanisms Through Which an Influential Early Childhood Program Boosted Adult Outcomes," *American Economic Review*, 2013, *103* (6), 2052–2086.

Huifeng, Y., J. Xiaotong, L. Guangke, H. Meng, and S. Nan, "Maternal Exposure to PM2.5 Affects Fetal Lung Development at Sensitive Windows," *Environmental Science & Technology*, 2020, *54* (1), 316–324.

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

Kearney, M. and R. Wilson, "Male Earnings, Marriageable Men, and Nonmarital Fertility: Evidence from the Fracking Boom," *Review of Economics and Statistics*, 2018, *100* (4), 678–690.

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

Koralek, D., "Reading Aloud with Children of All Ages," Technical Report, Reading is Fundamental 2014.

Lee, J. and L. Taylor, "Randomized Safety Inspections and Risk Exposure on the Job: Quasi-experimental Estimates of the Value of a Statistical Life," *American Economic Journal: Applied Economics*, 2019, *11* (4), 350–74.

Lovenheim, M., "The Effect of Liquid Housing Wealth on College Enrollment," *Journal of Labor Economics*, 2011, *29* (4), 741–771.

_ **and C. Lockwood Reynolds**, "The Effect of Housing Wealth on College Choice: Evidence from the Housing Boom," *Journal of Human Resources*, 2013, *14* (1), 3–37.

_ **and K.J. Mumford**, "Do Family Wealth Shocks Affect Fertility Choices? Evidence from the Housing Market," *Review of Economics and Statistics*, 2013, *95* (2), 464–475.

Lundberg, S., "Non-Cognitive Skills as Human Capital," *Mimeo*, 2017.

Massaro, D., "Reading Aloud to Children: Benefits and Implications for Acquiring Literacy Before Schooling Begins," *The American Journal of Psychology*, 2017, *130* (1), 63–72.

Mendelsohn, A., C. Brockmeyer, A. Weisleder, S. Berkule Johnson, A. Seery, C. Canfield, H. Huberman, and B. Dreyer, "Reading Aloud, Play, and Social-Emotional Development," *Pediatrics*, 2018, *141* (5).

Mohai, P., D. Pellow, and J.T. Roberts, "Environmental Justice," *Annual Review of Environment and Resources*, 2009, *34*, 405–430.

Murnane, R., J. Willett, Y. Duhaldeborde, and J. Tyler, "How Important Are the Cognitive Skills of Teenagers in Predicting Subsequent Earnings?," *Journal of Policy Analysis and Management*, 2000, *19* (4), 547–568.

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

Schaller, J., "Booms, Busts, and Fertility," *Journal of Human Resources*, 2016, *51* (1), 1–29.

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

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

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

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

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

Tables and Figures

Figure 1: The First-Stage Relationship between Nonattainment and First-Generation TSP Exposure (Second-Generation College Attendance Sample)

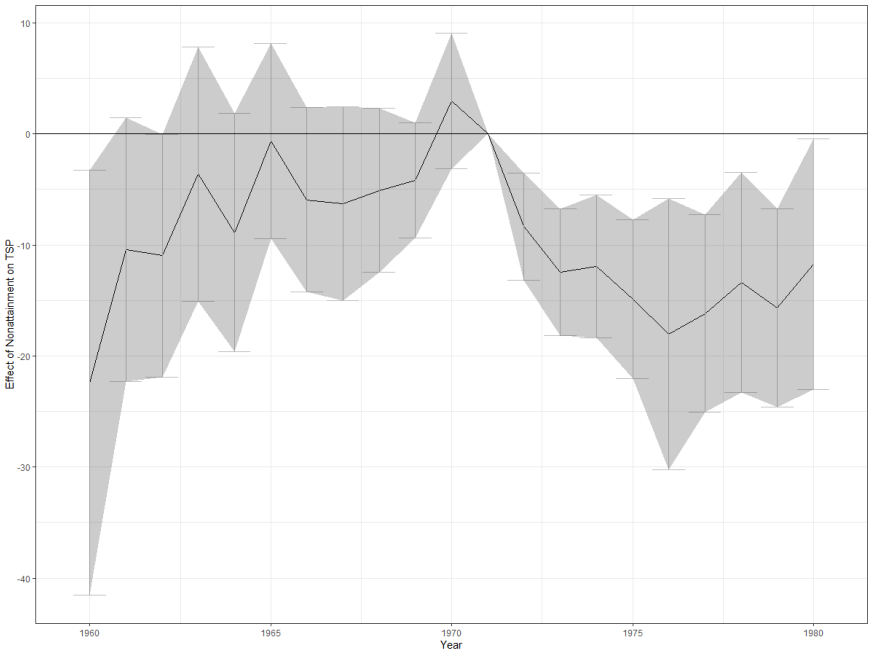

Table 1: The First-Stage Relationship between Nonattainment and First-Generation TSP Exposure (Second-Generation College Attendance Sample)

	First-Generation Prenatal TSP Exposure			
	(1)	(2)	(3)	(4)
Nonattainment × Post	-8.043***	-8.038***	-7.883***	-7.795***
	(2.231)	(2.232)	(2.218)	(2.213)
First-Gen Fixed Effects	County-of-birth, State-of-birth × Year, Birth Month			
First-Gen Controls	Yes	Yes	Yes	Yes
Second-Gen Controls	No	Yes	Yes	Yes
Second-Gen County FE?	No	No	Yes	Yes
Second-Gen SY FE?	No	No	No	Yes
Observations	238,000	238,000	234,000	234,000
Control Mean	79.27	79.27	79.31	79.31
First Stage F-Stat	12.99	12.97	12.63	12.41

Notes: Significance levels are indicated as * 0.10 ** 0.05 *** 0.01. Source: Decennial Census Short Form 2000 and 2010 and 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 nine month gestational period. Column 2-4 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-FY2020-CES010-015, CBDRB-FY2020-CES010-017, CBDRB-FY2020-CES010-029 and CBDRB-FY2020-CES010-031.

Table 2: The Relationship between Nonattainment and Pollution Exposure Over the Life Cycle

	(1) First-Gen Prenatal Exposure ($\mu g/m^3$)	(2) First-Gen Later-Life Exposure ($\mu g/m^3$)	(3) Second-Gen Prenatal Exposure ($\mu g/m^3$)	(4) Second-Gen Later-Life Exposure ($\mu g/m^3$)
Nonattainment × Post	-9.951***	0.985	1.316	4.467
	(2.137)	(0.831)	(3.239)	(3.124)
First-Gen Fixed Effects	County-of-birth, State-of-birth × Year, Birth Month			
First-Gen Prenatal Controls	Yes	Yes	Yes	Yes
Observations	4,773,000	2,551,000	377,000	219,000
Control Mean	79.27	39.84	53.91	39.2

Notes: Significance levels are indicated as * 0.10 ** 0.05 *** 0.01. Source: Census Numident, Decennial Census Short Form 2000 and 2010 and 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 nine month gestational period. Column 3-4 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-FY2020-CES010-015, CBDRB-FY2020-CES010-017, CBDRB-FY2020-CES010-029 and CBDRB-FY2020-CES010-031.

Table 3: The Effect of First-Generation Prenatal TSP Exposure on Second-Generation College Attendance

	Second-Generation College Attendance			
	(1)	(2)	(3)	(4)
Panel A: IV				
First-Gen Prenatal TSP ($10\mu g/m^3$)	−0.0353**	−0.0325**	−0.0364**	−0.0385**
	(0.0143)	(0.0138)	(0.0150)	(0.0156)
Panel B: Reduced Form				
Nonattainment × Post	0.0284***	0.0261***	0.0287***	0.0300***
	(0.0078)	(0.0078)	(0.0079)	(0.0079)
First-Gen Fixed Effects	County-of-birth, State-of-birth × Year, Birth Month			
First-Gen Controls	Yes	Yes	Yes	Yes
Second-Gen Controls	No	Yes	Yes	Yes
Second-Gen County FE?	No	No	Yes	Yes
Second-Gen SY FE?	No	No	No	Yes
Observations	238,000	238,000	234,000	234,000
Control Mean	0.529	0.529	0.528	0.528
First Stage F-Stat	12.99	12.97	12.63	12.41

Notes: Significance levels are indicated as * 0.10 ** 0.05 *** 0.01. Source: Census Numident, Decennial Census Short Form 2000 and 2010 and 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 nine month gestational period. Column 2-4 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-FY2020-CES010-015, CBDRB-FY2020-CES010-017, CBDRB-FY2020-CES010-029 and CBDRB-FY2020-CES010-031.

Table 4: The Effect of First-Generation Prenatal TSP Exposure on Second-Generation College Attendance: Adopted vs. Biological Children

	First-Gen Adopt	Second-Generation College Attendance			
	(1)	(2)	(3)	(4)	(5)
Panel A: IV					
First-Gen Prenatal TSP ($10\mu g/m^3$)	0.0014	-0.0353**	-0.0323**	-0.0357**	-0.03768**
	(0.0030)	(0.0143)	(0.0137)	(0.0149)	(0.0155)
First-Gen Prenatal TSP		-0.0046	-0.0073	-0.0164	-0.0204
× Second-Gen Adopted		(0.0259)	(0.0252)	(0.0256)	(0.0253)
Panel B: Reduced Form					
Nonattainment × Post	-0.0023	0.0283***	0.0258***	0.0277***	0.0289***
	(0.0026)	(0.0079)	(0.0080)	(0.0081)	(0.0080)
Nonattainment × Post		0.0068	0.0116	0.0299	0.0361
× Second-Gen Adopted		(0.0444)	(0.0431)	(0.0439)	(0.0426)
First-Gen Fixed Effects	County-of-birth, State-of-birth × Year, Birth Month				
First-Gen Controls	Yes	Yes	Yes	Yes	Yes
Second-Gen Controls	No	No	Yes	Yes	Yes
Second-Gen County FE?	No	No	No	Yes	Yes
Second-Gen SY FE?	No	No	No	No	Yes
Observations	4,773,000	238,000	238,000	234,000	234,000
Control Mean	0.11	0.529	0.529	0.528	0.528
First Stage F-Stat	20.02	6.50	6.49	6.4	6.25

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 nine month gestational period. 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-FY2020-CES010-015, CBDRB-FY2020-CES010-017, CBDRB-FY2020-CES010-029 and CBDRB-FY2020-CES010-031.

Table 5: The Effect of First-Generation Prenatal TSP Exposure on First-Generation Labor Market Outcomes

	(1) logEarnings	(2) Labor Force Participation	(3) Unemployed	(4) Public Assistance
Panel A: IV				
First-Gen Prenatal TSP ($10\mu g/m^3$)	-0.012**	0.0025	-0.0003	-0.0017
	(0.0061)	(0.0019)	(0.0009)	(0.0006
Panel B: Reduced Form				
Nonattainment × Post	0.013**	-0.0026	0.0003	0.0018
	(0.0060)	(0.0019)	(0.0009)	(0.0005)
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	4,382,000	6,477,000	6,477,000	9,060,000
Control Mean	$30,638	0.823	0.0456	0.006
First Stage F-Stat	19.86	18.8	18.8	18.93

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 nine month gestational period. Standard errors are clustered at the county of birth level. Approved for release by the Census DRB, authorization numbers CBDRB-FY2020-CES010-015, CBDRB-FY2020-CES010-017, CBDRB-FY2020-CES010-029 and CBDRB-FY2020-CES010-031.

Table 6: The Effect of First-Generation Prenatal TSP Exposure on First-Generation Family Structure Outcomes

	(1) Married	(2) Any Children	(3) # of Children	(4) Age at First Birth
Panel A: IV				
First-Gen Prenatal TSP ($10\mu g/m^3$)	0.0027	0.0037	0.0151	0.0046
	(0.0021)	(0.0045)	(0.0115)	(0.0327)
Panel B: Reduced Form				
Nonattainment × Post	-0.0028	-0.0038	-0.0157	-0.0049
	(0.0022)	(0.0047)	(0.0116)	(0.0339)
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	9,060,000	4,583,000	4,785,000	4,807,000
Control Mean	0.624	0.584	1.103	26.2
First Stage F-Stat	18.93	18.78	19.04	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 nine month gestational period. Standard errors are clustered at the county of birth level. Approved for release by the Census DRB, authorization numbers CBDRB-FY2020-CES010-015, CBDRB-FY2020-CES010-017, CBDRB-FY2020-CES010-029 and CBDRB-FY2020-CES010-031.

Table 7: The Effect of In-Utero TSP Exposure on Parental Time-Use

	(1) Reading to Kids (minutes)	(2) Educational Activities (minutes)	(3) Time Spent with Kids (minutes)
Panel A: IV			
First-Gen Prenatal TSP ($10\mu g/m^3$)	-1.459**	-0.531	-13.96
	(0.686)	(1.06)	(18.82)
Panel B: Reduced Form			
Nonattainment × Post	1.488**	0.541	14.24
	(0.624)	(1.063)	(19.58)
Fixed Effects	County-of-birth, State-of-birth × Year, Birth Month and Interview Day		
Individual Controls	Yes	Yes	Yes
County-level Controls	Yes	Yes	Yes
Observations	9,000	9,000	9,000
Control Mean	2.69	3.10	316.6
First Stage F-Stat	17.03	17.03	17.03

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 CBDRB-FY19-099, CBDRB-FY2020-CES010-001 and CBDRB-FY21-CES014-004.

Online Appendices—Not for Publication

Contents

A Main Results: Additional Results and Robustness Tests

In this appendix we present a series of additional results and robustness tests supporting our main result, documenting the effects of first-generation prenatal TSP exposure on second-generation college attendance.

A.1 Descriptive Statistics

Table A1: Descriptive Statistics Prior to the 1970 CAAA

	(1) Nonattainment County	(2) Attainment County	(3) Difference (2)-(1)
Cohort Demographics			
Female	0.430	0.425	0.005**
	(0.00157)	(0.00229)	(0.00278)
Black	0.139	0.105	0.034
	(0.0129)	(0.0174)	(0.0216)
White	0.671	0.694	-0.023
	(0.0185)	(0.0271)	(0.0328)
Hispanic	0.094	0.082	0.012
	(0.0181)	(0.0219)	(0.0284)
Other	0.010	0.023	-0.013*
	(0.0012)	(0.0067)	(0.0068)
Environmental Exposure			
Gestational Exposure (10 $\mu g/m^3$)	11.83	7.93	3.895***
	(0.3172)	(0.2464)	(0.4017)
Rainfall (days)	23	25.5	-2.505
	(1.22)	(1.17)	(1.69)
Avg. Daily Maximum Temperature	17.95	18.33	-0.3831
	(0.5595)	(0.5903)	(0.8133)
County Socio-Economic Characteristics			
County Population	1,628,000	458,000	1,170,000**
	(512,000)	(54,740)	(514,900)
Personal Income per Capita ($ Year)	4,245	4,003	242*
	(113.9)	(92.51)	(146.8)

Notes: Significance levels are indicated as * 0.10 ** 0.05 *** 0.01. Standard errors are clusterd at the parent's county of birth level. Source: Census Numident, Decennial Census Short Form 2000 and 2010, ACS 2005 through 2015. Approved for release by the Census DRB, authorization numbers CBDRB-FY2020-CES010-015, CBDRB-FY2020-CES010-017, CBDRB-FY2020-CES010-029 and CBDRB-FY2020-CES010-031.

A.2 Migration

Table A2: The Effect of First-Generation Prenatal TSP Exposure on Migration out of First-gen County of Birth

	(1) First-Gen Migration (Out of County)	(2) First-Gen Migration (Out of State)	(3) Second-Gen Migration (Out of County)	(4) Second-Gen Migration (Out of State)
Panel A: IV				
First-Gen Prenatal TSP ($10\mu g/m^3$)	0.0078**	0.0013	-0.0082	0.0017
	(0.0034)	(0.0027)	(0.0104)	(0.0076)
Panel B: Reduced Form				
Nonattainment × Post	-0.0078**	-0.0013	0.0068	-0.0014
	(0.0033)	(0.0026)	(0.0084)	(0.0063)
Fixed Effects	County-of-birth, State-of-birth × Year, Birth Month			
First-Gen Controls	Yes	Yes	Yes	Yes
Observations	4,773,000	4,773,000	373,000	373,000
Control Mean	0.718	0.386	0.573	0.315
First Stage F-Stat	21.68	21.68	14.92	14.92

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 and the Census MAFARF. 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 nine month gestational period. Column 3-4 contain additional second generation demographic characteristics including race/ethnicity, sex and a quadratic in age. Standard errors are clustered at the county of birth level. Approved for release by the Census DRB, authorization numbers CBDRB-FY2020-CES010-015, CBDRB-FY2020-CES010-017, CBDRB-FY2020-CES010-029 and CBDRB-FY2020-CES010-031.

A.3 The Effect of First-Generation Prenatal TSP Exposure on the likelihood of Second-Generation High School Completion

In addition to exploring the effects of parental prenatal pollution exposure on the likelihood of attending college we also explore the likelihood of dropping out of high school. Consistent with the college attendance effects we estimate a reduction in the likelihood of completing high school. The effect size is smaller, corresponding to a 0.5 percentage point reduction. However, high school dropouts are rarer events than attending college and so the effect size relative to the mean is similar (6%).

Table A3: The Effect of First-Generation Prenatal TSP Exposure on Second-Generation High School Completion

	Second-Generation High School Dropout			
	(1)	(2)	(3)	(4)
Panel A: IV				
First-Gen Prenatal TSP ($10\mu g/m^3$)	−0.0049*	−0.0045*	−0.0050*	−0.0050*
	(0.0028)	(0.0027)	(0.0028)	(0.0028)
Panel B: Reduced Form				
Nonattainment × Post	0.0041**	0.0037*	0.0041**	0.0041**
	(0.0019)	(0.0019)	(0.0019)	(0.0019)
First-Gen Fixed Effects	County-of-birth, State-of-birth × Year, and Birth Month			
First-Gen Controls	Yes	Yes	Yes	Yes
Second-Gen Controls	No	Yes	Yes	Yes
Second-Gen County FE?	No	No	Yes	Yes
Second-Gen SY FE?	No	No	No	Yes
Observations	942,000	942,000	924,000	924,0000
Control Mean	0.067	0.067	0.067	0.067
First Stage F-Stat	13.8	13.8	13.78	13.77

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 nine month gestational period. Column 2-4 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-FY2020-CES010-015, CBDRB-FY2020-CES010-017, CBDRB-FY2020-CES010-029 and CBDRB-FY2020-CES010-031.

A.4 Alternate Sample Restrictions

In our main analysis we focus on a sample that is based on certain parent-child links, include all births between 1960 and 1980, and define "treatment" to be assigned if any parent-child link was affected. The following results show that our findings are robust to using a more narrow window of births around the 1970 CAAA, to including probabilistic parent-child links, and to restricting "treatment definitions" based on maternal or paternal exposure.

A.4.1 Different Treatment Windows

Here we present estimates of the effects of first-generation TSP exposure on second-generation college attendance, restricting the sample to two alternative treatment windows: 1969-1975 and 1969-1980.

Table A4: The Effect of First-Generation Prenatal TSP Exposure on Second-Generation College Attendance (1969-1975 Window)

	Second-Generation College Attendance			
	(1)	(2)	(3)	(4)
Panel A: IV				
First-Gen Prenatal TSP ($10\mu g/m^3$)	-0.0253**	-0.0229**	-0.0251**	-0.0260**
	(0.0116)	(0.0112)	(0.0120)	(0.0125)
Panel B: Reduced Form				
Nonattainment × Post	0.0289***	0.0262**	0.0279***	0.0287***
	(0.0104)	(0.0105)	(0.0107)	(0.0109)
First-Gen Fixed Effects	County-of-birth, State-of-birth × Year, and Birth Month			
First-Gen Controls	Yes	Yes	Yes	Yes
Second-Gen Controls	No	Yes	Yes	Yes
Second-Gen County FE?	No	No	Yes	Yes
Second-Gen SY FE?	No	No	No	Yes
Observations	71,000	71,000	69,500	69,500
Control Mean	0.423	0.423	0.423	0.423
First Stage F-Stat	12.85	12.83	13.24	13.00

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 nine month gestational period. Column 2-4 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-FY2020-CES010-015, CBDRB-FY2020-CES010-017, CBDRB-FY2020-CES010-029 and CBDRB-FY2020-CES010-031.

Table A5: The Effect of First-Generation Prenatal TSP Exposure on Second-Generation College Attendance (1969-1980 Window)

	Second-Generation College Attendance			
	(1)	(2)	(3)	(4)
Panel A: IV				
First-Gen Prenatal TSP ($10\mu g/m^3$)	-0.0353**	-0.0325**	-0.0364**	-0.0385**
	(0.0143)	(0.0138)	(0.0150)	(0.0156)
Panel B: Reduced Form				
Nonattainment × Post	0.0278	0.0254	0.0257	0.0265
	(0.0099)	(0.0101)	(0.0100)	(0.0103)
First-Gen Fixed Effects	County-of-birth, State-of-birth × Year, and Birth Month			
First-Gen Controls	Yes	Yes	Yes	Yes
Second-Gen Controls	No	Yes	Yes	Yes
Second-Gen County FE?	No	No	Yes	Yes
Second-Gen SY FE?	No	No	No	Yes
Observations	78,500	78,500	77,000	77,000
Control Mean	0.418	0.418	0.418	0.418
First Stage F-Stat	12.99	12.97	12.63	12.41

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 nine month gestational period. Column 2-4 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-FY2020-CES010-015, CBDRB-FY2020-CES010-017, CBDRB-FY2020-CES010-029 and CBDRB-FY2020-CES010-031.

A.4.2 Including Probabilistic Links

Here we present results on the effects of first-generation TSP exposure on second-generation college attendance, using an expanded sample that incorporates probabilistic parent-child links. This increases our sample size from 234,000 to 357,000. We estimate similar effects, although the magnitude is slightly smaller consistent with the introduction of measurement error that arises from any false links.

Table A6: The Effect of First-Generation Prenatal TSP Exposure on Second-Generation College Attendance (including Probabilistic Parent-Child Links)

	Second-Generation College Attendance			
	(1)	(2)	(3)	(4)
Panel A: IV				
First-Gen Prenatal TSP ($10\mu g/m^3$)	-0.0239**	-0.0216**	-0.0248**	-0.0259**
	(0.0106)	(0.0102)	(0.0109)	(0.0112)
Panel B: Reduced Form				
Nonattainment × Post	0.0197***	0.0178***	0.0203***	0.0210***
	(0.0066)	(0.0066)	(0.0067)	(0.0066)
First-Gen Fixed Effects	County-of-birth, State-of-birth × Year, and Birth Month			
First-Gen Controls	Yes	Yes	Yes	Yes
Second-Gen Controls	No	Yes	Yes	Yes
Second-Gen County FE?	No	No	Yes	Yes
Second-Gen SY FE?	No	No	No	Yes
Observations	364,000	364,000	357,000	357,000
Control Mean	0.506	0.506	0.506	0.506
First Stage F-Stat	14.51	14.49	14.27	14.04

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 nine month gestational period. Column 2-4 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-FY2020-CES010-015, CBDRB-FY2020-CES010-017, CBDRB-FY2020-CES010-029 and CBDRB-FY2020-CES010-031.

A.4.3 Maternal vs. Paternal Exposure

In this section we explore the effects of first-generation TSP exposure on second-generation college attendance restricting the treatment group to: cases in which mother's were affected; cases in which father's were affected; cases in which both the mother and father were affected. We estimate slightly larger effects when both the mother and father were affected.

Table A7: The Effect of First-Generation Prenatal TSP Exposure on Second-Generation College Attendance (Mother Exposed)

	Second-Generation College Attendance			
	(1)	(2)	(3)	(4)
Panel A: IV				
First-Gen Prenatal TSP ($10\mu g/m^3$)	−0.0440***	−0.0395**	−0.0419**	−0.0423**
	(0.0170)	(0.0161)	(0.0172)	(0.0175)
Panel B: Reduced Form				
Nonattainment × Post	0.0344***	0.0309***	0.0320***	0.0320***
	(0.0084)	(0.0086)	(0.0087)	(0.0086)
First-Gen Fixed Effects	County-of-birth, State-of-birth × Year, and Birth Month			
First-Gen Controls	Yes	Yes	Yes	Yes
Second-Gen Controls	No	Yes	Yes	Yes
Second-Gen County FE?	No	No	Yes	Yes
Second-Gen SY FE?	No	No	No	Yes
Observations	227,000	227000	223,000	223,000
Control Mean	0.5	0.5	0.5	0.5
First Stage F-Stat	12.5	12.48	12.09	11.85

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. The sample is restricted to compare the outcomes of children who's mother's, were affected (but not father's) to children where neither mother or father were affected. 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 nine month gestational period. Column 2-4 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-FY2020-CES010-015, CBDRB-FY2020-CES010-017, CBDRB-FY2020-CES010-029 and CBDRB-FY2020-CES010-031.

Table A8: The Effect of First-Generation Prenatal TSP Exposure on Second-Generation College Attendance (Father Exposed)

	Second-Generation College Attendance			
	(1)	(2)	(3)	(4)
Panel A: IV				
First-Gen Prenatal TSP ($10\mu g/m^3$)	-0.0410**	-0.0376**	-0.0424**	-0.0448**
	(0.0181)	(0.0176)	(0.0194)	(0.0202)
Panel B: Reduced Form				
Nonattainment × Post	0.0303***	0.0278***	0.0306***	0.0321***
	(0.0092)	(0.0093)	(0.0096)	(0.0097)
First-Gen Fixed Effects	County-of-birth, State-of-birth × Year, Birth Month			
First-Gen Controls	Yes	Yes	Yes	Yes
Second-Gen Controls	No	Yes	Yes	Yes
Second-Gen County FE?	No	No	Yes	Yes
Second-Gen SY FE?	No	No	No	Yes
Observations	219,000	219,000	215,000	215,000
Control Mean	0.499	0.499	0.499	0.499
First Stage F-Stat	11.18	11.15	10.67	10.51

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. The sample is restricted to compare the outcomes of children who's father's, were affected (but not mother's) to children where neither mother or father were affected. 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 nine month gestational period. Column 2-4 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-FY2020-CES010-015, CBDRB-FY2020-CES010-017, CBDRB-FY2020-CES010-029 and CBDRB-FY2020-CES010-031.

Table A9: The Effect of First-Generation Prenatal TSP Exposure on Second-Generation College Attendance (Both Mother and Father Exposed)

	Second-Generation College Attendance			
	(1)	(2)	(3)	(4)
Panel A: IV				
First-Gen Prenatal TSP ($10\mu g/m^3$)	-0.0665**	-0.0558**	-0.0568**	-0.0574**
	(0.0268)	(0.0243)	(0.0257)	(0.0262)
Panel B: Reduced Form				
Nonattainment × Post	0.0467***	0.0391***	0.0386***	0.0386***
	(0.0108)	(0.0110)	(0.0113)	(0.0113)
First-Gen Fixed Effects	County-of-birth, State-of-birth × Year, Birth Month			
First-Gen Controls	Yes	Yes	Yes	Yes
Second-Gen Controls	No	Yes	Yes	Yes
Second-Gen County FE?	No	No	Yes	Yes
Second-Gen SY FE?	No	No	No	Yes
Observations	208,000	208,000	204,000	204,000
Control Mean	0.502	0.502	0.502	0.502
First Stage F-Stat	9.86	9.82	9.25	9.06

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. The sample is restricted to compare the outcomes of children where both parents were affected to children where neither parent was affected. 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 nine month gestational period. Column 2-4 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-FY2020-CES010-015, CBDRB-FY2020-CES010-017, CBDRB-FY2020-CES010-029 and CBDRB-FY2020-CES010-031.

A.5 Regression Discontinuity Design Approach

As there is a sharp cutoff in assignment to nonattainment status, we also test the robustness of our findings to using a regression discontinuity design.

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, including a linear spline for pre-1970 CAAA average TSP concentrations. 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:

$$\begin{aligned} Outcome_{i,j,c,t} &= \alpha_0 + \alpha_1(Nonattainment_{c,1970} \times 1[\tau > 1971]) \\ &\quad + \alpha_2 Dist_{c,1970} + \alpha_3 Dist_{c,1970} \times 1[TSP_c > 75] + \gamma X_j' + \delta X_c' t + \epsilon_{i,j,c,t} \end{aligned}$$

We estimate these RD regressions with bandwidths varying from $25 \mu g/m^3$ to 150 $\mu g/m^3$. Our main specification uses the largest analysis window (1960-1980), certain parent-child links, and does not include second-generation controls. We estimate similar effects to our difference-in-differences specification. Results are robust to including probabilistic parent-child links, to narrowing the analysis window (1969-1980 or 1969-1974), and to including second-generation controls.

Table A10: The Effect of First-Generation Prenatal TSP Exposure on Second-Generation College Attendance (RDD)

	Second-Generation College Attendance			
	(1)	(2)	(3)	(4)
Panel A: Baseline specification				
Nonattainment	0.030*	0.014	0.018*	0.023**
	(0.017)	(0.011)	(0.010)	(0.010)
Observations	99,500	197,000	291,000	301,000
Panel B: Second-Gen Controls				
Nonattainment	0.037**	0.020*	0.024**	0.030***
	(0.016)	(0.011)	(0.010)	(0.010)
Observations	91,000	180,000	266,000	275,000
Panel C: Probabilistic Links				
Nonattainment	0.027**	0.016*	0.026***	0.030***
	(0.012)	(0.009)	(0.008)	(0.008)
Observations	139,000	275,000	406,000	420,000
Panel D: 1969-1980 Window				
Nonattainment	0.052**	0.024	0.027**	0.030**
	(0.024)	(0.016)	(0.013)	(0.013)
Observations	24,500	48,500	71,500	73,500
Bandwidth	25	50	100	150

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 nine month gestational period. Column 2-4 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-FY2020-CES010-015, CBDRB-FY2020-CES010-017, CBDRB-FY2020-CES010-029 and CBDRB-FY2020-CES010-031.

B Exploring Mechanisms: Additional Results and Robustness Tests

B.1 Step-Children vs. Biological Children

Table B1: The Effect of First-Generation Prenatal TSP Exposure on Second-Generation College Attendance: Step-Children vs. Biological Children

	Has Step-Child	Second-Generation College Attendance			
	(1)	(2)	(3)	(4)	(5)
Panel A: IV					
First-Gen Prenatal TSP ($10\mu g/m^3$)	0.0096**	-0.0346**	-0.0326**	-0.0359**	-0.0379**
	(0.0048)	(0.0160)	(0.0156)	(0.0168)	(0.0174)
First-Gen Prenatal TSP		-0.0054	-0.0043	-0.0056	-0.0058
× Second-Gen Step-Child		(0.0097)	(0.0095)	(0.0097)	(0.0097)
Panel B: Reduced Form					
Nonattainment × Post	-0.0091**	0.0256***	0.0241***	0.0259***	0.0271***
	(0.0042)	(0.0086)	(0.0087)	(0.0087)	(0.0087)
Nonattainment × Post		0.0170	0.0147	0.0180	0.0190
× Second-Gen Step-Child		(0.0164)	(0.0161)	(0.0163)	(0.0163)
First-Gen Fixed Effects	County-of-birth, State-of-birth × Year, Birth Month				
First-Gen Controls	Yes	Yes	Yes	Yes	Yes
Second-Gen Controls	No	No	Yes	Yes	Yes
Second-Gen County FE?	No	No	No	Yes	Yes
Second-Gen SY FE?	No	No	No	No	Yes
Observations	2,493,000	238,000	238,000	234,000	234,000
Control Mean	0.29	0.5178	0.5178	0.5178	0.5178
First Stage F-Stat	21.3	10.31	10.33	10.54	10.62

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 nine month gestational period. Column 2-4 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-FY2020-CES010-015, CBDRB-FY2020-CES010-017, CBDRB-FY2020-CES010-029 and CBDRB-FY2020-CES010-031.

B.2 Fertility Results by Sex

Table B2: The Effect of First-Generation Prenatal TSP Exposure on First-Generation Family Structure Outcomes (Women)

	(1) Married	(2) # of Children	(3) Age at First Birth
Panel A: IV			
First-Gen Prenatal TSP ($10\mu g/m^3$)	-0.00035	0.0075	0.0013
	(0.0026)	(0.013)	(0.045)
Panel B: Reduced Form			
Nonattainment × Post	0.00036	-0.0076	-0.0014
	(0.0027)	(0.013)	(0.046)
Fixed Effects	County-of-birth, State-of-birth × Year, Birth Month		
Individual Controls	Yes	Yes	Yes
County-level Controls	Yes	Yes	Yes
Observations	4,624,000	2,445,000	2,710,000
Control Mean	0.63	1.14	25.4
First Stage F-Stat	18.75	18.9	18.79

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 nine month gestational period. Standard errors are clustered at the county of birth level. Approved for release by the Census DRB, authorization numbers CBDRB-FY2020-CES010-015, CBDRB-FY2020-CES010-017, CBDRB-FY2020-CES010-029 and CBDRB-FY2020-CES010-031.

Table B3: The Effect of First-Generation Prenatal TSP Exposure on First-Generation Family Structure Outcomes (Men)

	(1) Married	(2) # of Children	(3) Teen Pregnancy
Panel A: IV			
First-Gen Prenatal TSP ($10\mu g/m^3$)	0.006*	0.025*	0.015
	(0.0032)	(0.013)	(0.042)
Panel B: Reduced Form			
Nonattainment × Post	-0.0063**	-0.027**	-0.016
	(0.0029)	(0.013)	(0.044)
Fixed Effects	County-of-birth, State-of-birth × Year, Birth Month		
Individual Controls	Yes	Yes	Yes
County-level Controls	Yes	Yes	Yes
Observations	4,437,000	2,330,000	2,096,000
Control Mean	0.62	1.06	27.25
First Stage F-Stat	19.06	18.98	19.46

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 nine month gestational period. Standard errors are clustered at the county of birth level. Approved for release by the Census DRB, authorization numbers CBDRB-FY2020-CES010-015, CBDRB-FY2020-CES010-017, CBDRB-FY2020-CES010-029 and CBDRB-FY2020-CES010-031.

B.3 First-Generation Educational Outcomes

Table B4: The Effect of First-Generation Prenatal TSP Exposure on First-Generation Educational Outcomes

	(1) Years of Schooling	(2) High School Dropout	(3) Attended College
Panel A: IV			
First-Gen Prenatal TSP ($10\mu g/m^3$)	-0.0103	-0.0008	-0.0003
	(0.0170)	(0.0012)	(0.0029)
Panel B: Reduced Form			
Nonattainment × Post	0.0108	0.0008	0.0003
	(0.0175)	(0.0012)	(0.0031)
Fixed Effects	County-of-birth, State-of-birth × Year, Birth Month		
Individual Controls	Yes	Yes	Yes
County-level Controls	Yes	Yes	Yes
Observations	6,485,000	6,485,000	6,485,000
Control Mean	14.84	0.05477	0.6932
First Stage F-Stat	18.81	18.81	18.81

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 nine month gestational period. Standard errors are clustered at the county of birth level. Approved for release by the Census DRB, authorization numbers CBDRB-FY2020-CES010-015, CBDRB-FY2020-CES010-017, CBDRB-FY2020-CES010-029 and CBDRB-FY2020-CES010-031.

B.4 Assortative Matching

Table B5: The Effect of First-Generation on Partner Characteristics

	(1) Partner Treated	(2) Both Employed	(3) Both College Educated	(4) Same Income Quintile
Panel A: IV				
First-Gen Prenatal TSP ($10\mu g/m^3$)	-0.64***	-0.0034	-0.0022	-0.0070*
	(0.15)	(0.0026)	(0.0028)	(0.0042)
Panel B: Reduced Form				
Nonattainment × Post	0.66***	0.0034	0.0022	0.0070*
	(0.0044)	(0.0025)	(0.0027)	(0.0041)
Fixed Effects	County-of-birth, State-of-birth × Year, Birth Month			
First-Gen Controls	Yes	Yes	Yes	Yes
Observations	2,244,000	2,241,000	2,243,000	1,665,000
Control Mean	0.203	0.690	0.289	0.681
First Stage F-Stat	18.4	20.91	20.94	21.47

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 nine month gestational period. Standard errors are clustered at the county of birth level. Approved for release by the Census DRB, authorization numbers CBDRB-FY2020-CES010-015, CBDRB-FY2020-CES010-017, CBDRB-FY2020-CES010-029 and CBDRB-FY2020-CES010-031.

C Data Appendix

C.1 The Census Bureau's 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 Census, survey and administrative data to identify parent-child 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.[16] Administrative 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,

[16]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.

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.

C.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).

C.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).

C.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 the American Community Survey, which has been conducted monthly since 2001, as our main survey source. The Current Population Survey Annual Social and Economic Supplement (CPS ASEC), which is conducted every March is another potential source of parent-child links. However, 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), and thus provides minimal additional information relative to the decennial Census and ACS.

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.

C.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. Our main estimating sample uses only the certain links found in the Census and ACS data. In robustness exercises, we additionally include probabilistic links.

C.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).

C.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 (described in detail in (Isen et al., 2017)). 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 (2017). 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.

C.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 particles (TSP), defined as the density of particulates less than 50 microns, measured in units $\mu g/m^3$.[17] 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 consistently monitored, the expansion of the monitor network may result in systematic measurement error—average county exposure will be more precisely measured with more

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

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).

C.3.3 Nonattainment Designations

Our empirical strategy relies on 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. This is because the EPA designated entire air quality regions as in nonattainment, not individual counties, when setting standards authorized under CAAA 1970. In practice, however, the way that these nonattainment designations were applied bound at the county level: the EPA considered an air quality district to be in nonattainment if any of the monitors in the region were in nonattainment. Thus in practice counties with monitors in attainment were not subject to the same regulations as counties that were in nonattainment. Since the TSP air quality standards are known, 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, it is possible to reconstruct which counties would have been in Nonattainment.

Nonattainment of the primary air quality standard for TSP set in CAA 1970 occurs if either a) the 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.

C.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 the National Oceanic and Atmospheric Administration (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.

C.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).