Every Breath You Take — Every Dollar You'll Make: The Long-Term Consequences of the Clean Air Act of 1970*

Adam Isen
Office of Tax Analysis
Department of the Treasury
adam.isen@treasury.gov

Maya Rossin-Slater
University of California,
Santa Barbara
maya.rossin-slater@ucsb.edu

W. Reed Walker
University of California,
Berkeley, NBER and IZA
rwalker@berkeley.edu

September 2015

Abstract

This paper examines the long-term impacts of early childhood exposure to air pollution on adult outcomes using U.S. administrative data. We exploit changes in air pollution driven by the 1970 Clean Air Act to analyze the difference in outcomes between cohorts born in counties before and after large improvements in air pollution relative to those same cohorts born in counties that had no improvements. We find a significant relationship between pollution exposure in the year of birth and later life outcomes. A higher pollution level in the year of birth is associated with lower labor force participation and lower earnings at age 30.

JEL: H40, H51, I12, I14, J17, J18, J31, Q51, Q53, Q58

*This paper has been previously circulated under the title "Does Improved Air Quality at Birth Lead to Better Long-Term Outcomes? Evidence from the Clean Air Act of 1970". We would like to thank Doug Almond, Michael Anderson, David Card, Janet Currie, Lucas Davis, Olivier Deschenes, Will Dow, Ilyana Kuziemko, Matt Neidell, Yona Rubinstein, four anonymous referees, as well as seminar participants at Columbia, Georgia State, Texas A&M, University of Calgary, UC-Berkeley, NBER Summer Institute, Census, IZA, the Robert Wood Johnson Foundation, the AEA meetings, the UCSB Occasional Workshop, and the 4th Annual All-UC Conference on Energy and Environmental Economics for valuable comments and suggestions. We also thank David Silver for helpful research assistance. All results have been reviewed to ensure that no confidential information is disclosed. This research uses data from the Census Bureau's Longitudinal Employer Household Dynamics Program, which was partially supported by the following NSF Grants SES-9978093, SES-0339191 and ITR-0427889; NIA Grant AG018854; and grants from the Alfred P. Sloan Foundation. Isen acknowledges additional support from the Institute of Education Sciences, U.S. Department of Education, through Grant R305B090015 of the U.S. Department of Education. Walker acknowledges additional support from the Robert Wood Johnson Foundation and the University of California Center for Energy and Environmental Economics. Research results and conclusions expressed are those of the authors and do not necessarily reflect the views of the Census Bureau, the Robert Wood Johnson Foundation, or the U.S. Department of the Treasury.

1 Introduction

The desire to protect human health and welfare motivates much of modern environmental regulation.[1] While there is growing evidence in both epidemiology and economics pointing to the contemporaneous influences of ambient air pollution on population health and other measures of welfare (see Graff Zivin and Neidell, 2013 for a recent review), there is a relative dearth of empirical evidence on the long-run and cumulative impacts of environmental toxins. We view this gap in the literature as particularly important because research has suggested a critical link between population health and wealth throughout the lifecycle (Currie, 2009). Therefore, contemporaneous measures of the dose-response relationship between environmental conditions and health outcomes may substantially underestimate the total welfare impact of environmental toxins.

This paper provides some of the first quasi-experimental evidence linking early-life environmental exposure to adult measures of well-being. To study this topic, we leverage a policy experiment in the early 1970s that generated large reductions in ambient pollution levels in hundreds of counties in the United States. We then examine whether cohorts that were born just before and just after these large changes in air pollution exhibit persistent differences in outcomes measured 30 years after birth. Since our comparison group includes individuals who were born in affected counties in the few years before the policy went into effect, our empirical design effectively isolates any *additional* impacts of exposure to cleaner air in very early childhood relative to such exposure at slightly older ages. Our focus on the early life stage is motivated by emerging evidence on the fetal origins of adult outcomes (Barker, 1990; Gluckman and Hanson, 2004; Almond and Currie, 2011), combined with the mounting evidence on the particularly severe impacts of pollution on infant and fetal health.[2]

We combine this policy experiment with newly available administrative data from the U.S. Census Bureau's Longitudinal Employer Household Dynamics (LEHD) file that allows us to observe adult outcomes linked to location and exact date of birth for 5.7 million individuals born around the time of the policy experiment. We focus on measures of labor market performance around age 30 that broadly encompass (i) changes to cognitive and non-cognitive skill formation that may have been "imprinted" in early childhood, (ii) any persistent health effects attributable to early-life air pollution exposure, and (iii) any reinforcing or compensatory parental investments.[3] As such, our outcomes represent quantifiable summary measures that may be particularly relevant for cost-benefit calculations

[1] See, for example, the Environmental Protection Agency's (EPA) mission statement at: http://www2.epa.gov/aboutepa/our-mission-and-what-we-do (accessed on June 13, 2013).

[2] The "fetal origins hypothesis," originally put forth by British epidemiologist David J. Barker, argues that poor nutrition *in-utero* "programs" the fetus to have metabolic characteristics that can lead to future disease in adulthood. For recent evidence on the link between pollution and infant/fetal health, see, for example: Chay and Greenstone, 2003a; Chay and Greenstone, 2003b; Currie and Neidell, 2005; Currie *et al.*, 2009b; Currie and Walker, 2011; Sanders and Stoecker, 2015.

[3] See Becker and Tomes (1976) for economic theory regarding parental responses to initial endowments. A number of recent studies have provided empirical evidence on parental responses to early life health (see, e.g., Adhvaryu and Nyshadham, 2011; Aizer and Cunha, 2010; Datar *et al.*, 2010; Del Bono *et al.*, 2012; Almond and Mazumder, 2012; Bharadwaj *et al.*, 2013a; Conti *et al.*, Forthcoming). Also, see Gelber and Isen (2013) for some related empirical evidence on complementarity in schooling and parental investment. Finally, Heckman and Mosso (2014) provide a detailed discussion of how cognitive and non-cognitive skills, health capital, and parental investments all interact in a model of human development.

in environmental policy design.

The policy experiment in the paper stems from the introduction of the 1970 Clean Air Act Amendments (CAAA), which imposed county-level restrictions on the maximum-allowable concentrations of total suspended particulates (TSP). As a result, counties that exceeded these new restrictions (nonattainment counties) were forced to reduce their TSP concentrations, while counties that had air pollution levels below the regulatory ceiling (attainment counties) were not legally required to change their TSP emissions. This legislation induced substantial variation in county-level pollution changes during the 1970s that has been previously used to study the effects of air pollution on infant mortality (Chay and Greenstone, 2003a), adult mortality (Chay *et al.*, 2003), and fetal mortality (Sanders and Stoecker, 2015). We use this variation to estimate whether cohorts exposed to lower levels of ambient air pollution *in-utero* and in the first year of life exhibit improved labor market outcomes measured some 30 years later. Our baseline empirical specification compares cohorts of individuals born just before and after the mandated improvements in air quality in nonattainment counties, using cohorts born in attainment counties as a counterfactual control group. While nonattainment status is not randomly assigned, we show that observable characteristics of nonattainment and attainment counties in the years prior to regulation are similar in both levels and, more importantly, trends.

Our results suggest that county-level air pollution in an individual's year of birth has a statistically significant and economically meaningful impact on labor market outcomes measured around age 30. We first show that the CAAA led to an over 10 percent reduction in ambient TSP levels in nonattainment counties in the three years after the regulation went into effect. We then show that this regulation-induced reduction in air pollution is associated with a 0.7 percent increase in the annual number of quarters worked and a 1 percent increase in mean annual earnings for affected cohorts. Assuming a constant earnings effect over the lifecycle, our results suggest that the cumulative lifetime income gain is approximately $4,300 in present value terms (using a 5% annual discount rate). This calculation implies that the present discounted total wage bill attributable to the improvements in early life air quality amounts to about $6.5 billion for each affected cohort. We view these estimates as lower bounds on the true value due to several potential sources of bias that would attenuate our baseline estimates (which we discuss later in the text). Nevertheless, our estimates suggest that the long-run welfare costs of exposure to environmental toxins as measured by lifetime earnings losses may be as large or larger than the monetized costs of death based on short-run impacts on infant mortality examined in previous research (e.g., Chay and Greenstone (2003a)).

This paper provides three primary contributions: First, prior literature estimating the health effects of the CAAA typically focuses on contemporaneous changes in infant health. While infant mortality is an important outcome to study, it reflects some of the most severe consequences of adverse environmental conditions. There may be other consequences for individuals who survive, and, as human capital is an engine for long-run economic growth (Romer, 1986; Schultz, 1961), in aggregate these effects may be larger and far more long-lasting than those associated with infant mortality gains (Graff Zivin and Neidell, 2013). Although there is some evidence of a contemporaneous relationship

between pollution and economic outcomes,[4] there is very little work that examines how the short-run benefits of environmental policy may persist in the long run.[5]

Second, we provide additional quasi-experimental empirical support for the theory of fetal origins and early-life determinants of long-run outcomes. A substantial literature has documented a strong relationship between markers of early-life health (such as birth weight and the presence of chronic conditions) and adult outcomes (see, e.g.: Black *et al.* (2007); Currie and Moretti (2007); Oreopoulos *et al.* (2008); Currie *et al.* (2010); Bharadwaj *et al.* (2013b)). A related line of research has studied the consequences of early-life shocks to health, mostly using variation from rare natural disasters, disease outbreaks, or famines, which are difficult to forecast or protect against (Almond (2006); Almond *et al.* (2009); Almond *et al.* (2010)). In contrast, we examine the long-run returns to environmental regulation, an intervention over which policy-makers have direct control. The dose-response relationship between ambient air pollution and long-run labor market performance is an important policy parameter for which we have very few estimates.

Third, this paper introduces a new resource for studying long-term outcomes in the United States: the Longitudinal Employer-Household Dynamic Files from the U.S. Census Bureau. Previous work focusing on long-run implications of early-life interventions in the United States has been typically challenged by the fact that very few publicly available datasets contain detailed information on birth location linked to long-run outcomes (the few that do are of limited use in this context because of small sample sizes).[6] In contrast, our administrative earnings data contain the near-universe of the employed workforce, with precise information on both location and date of birth.

While the LEHD has clear advantages over existing survey data sets, it has some limitations that bear mention. First, earnings records are only available in certain states and in certain years, and our baseline analysis uses data from 24 states, which are continuously part of the LEHD over 1998-2007. Employment in these states accounts for nearly two-thirds of the U.S. non-farm workforce. Second,

[4]For example, Hanna and Oliva (2011) examine labor supply, while Graff-Zivin and Neidell (2013) study labor productivity. Studies also show that contemporaneous pollution exposure can affect human capital accumulation by increasing school absenteeism (Ransom and Pope, 1992; Gilliland *et al.*, 2001; Currie *et al.*, 2009a), and impairing cognitive performance on high-stakes tests (Lavy *et al.*, 2012). There is also a possibility that pollution can affect adult income if parents have to forego work to take care of asthmatic children (Currie *et al.*, 2009a).

[5]Within the United States, we are only aware of two papers that study these questions, although they focus on non-labor market outcomes. Sanders (2012) analyzes the relationship between early-life air pollution and high school test scores in Texas, while Reyes (2007) examines the effects of early-life lead exposure on young adult crime. However, an important limitation of both studies is the lack of information on place of birth. As a result, Sanders (2012) effectively assigns birth location based on county of high school attendance, while Reyes (2007) assigns exposure based on state of crime occurrence around age 20. These analyses may therefore be affected by bias from endogenous mobility responses and measurement error. Nonetheless, these studies suggest that there might be an earnings effect of pollution. Outside the U.S., Bharadwaj *et al.* (2014) study the impacts of fetal exposure to air pollution on 4th grade test scores in Santiago, Chile in a sibling fixed effects design, finding negative impacts on math and language scores. Additionally, two recent studies have estimated the impacts of early-life lead exposure on adult outcomes in Sweden (Nilsson, 2009) and Chile (Rau *et al.*, 2013). However, Bharadwaj *et al.* (2014) and Rau *et al.* (2013) are limited in their ability to directly observe labor market outcomes. Nilsson (2009) and Rau *et al.* (2013) focus on a very different and far more toxic pollutant (lead) in a context outside of the United States.

[6]The restricted version of the Panel Study of Income Dynamics (PSID) is the best currently available dataset that gives information on location and date of birth linked to long-run earnings (Johnson and Schoeni, 2007; Hoynes *et al.*, 2012; Johnson, 2011).

as in most administrative earnings data sets, we cannot discern between missing earnings records that occur because of non-employment and those due to employment in a state outside our LEHD sample. We take a number of steps in the paper to ensure that our results are not driven by differential sample attrition. For example, we show that cross-state mobility within our 24 sample states is not correlated with treatment status (i.e., cohorts born into nonattainment counties after CAAA implementation are no more or less likely to move away from their home state relative to the comparison cohorts). We also present results using workers with non-zero earnings, and we show that the effects, while attenuated, are similar.

The rest of the paper proceeds as follows: Section 2 presents a basic conceptual framework to help guide the empirical analysis. Section 3 provides a brief overview of the CAAA and related literature. Section 4 provides a description of the data used in the analysis, with a more complete discussion found in Appendix C. Section 5 outlines the various econometric models used, and Section 6 discusses the results of those models. Sections 7 and 8 discusses the implications of our findings and concludes, respectively.

2 Conceptual Framework

How might early-life exposure to ambient air pollution affect adult outcomes? In this paper, we focus on early-life exposure to TSP, which is the type of pollution regulated by the EPA at the time of the 1970 CAAA. TSPs include all suspended airborne solid or liquid particles found in the air that are smaller than 100 micrometers in size. TSPs enter the atmosphere both from human sources (such as motor vehicles and industrial activities) and natural sources (such as dust, dirt, and pollen). Some of these particles are large enough to be seen as soot or smoke, while others are so small that they can only be detected with an electron microscope.

In terms of damage to human health, bigger particulates are less harmful than smaller ones. The larger and heavier particles settle to the ground quickly and are less likely to be inhaled by humans relative to smaller particles. When they are inhaled, the larger particles settle in the nose and throat, and can usually be eliminated from the body through sneezing and coughing. In contrast, smaller particulate matter (e.g., particles less than 10 micrometers in size) can remain in the air for days or weeks, and once inhaled, can penetrate deep into the lung system.[7] These smaller particles collect in tiny air sacs in the lungs (alveoli) where oxygen enters the bloodstream.

Inhaled TSPs can affect respiratory function and lung development. Moreover, since particulates can be transferred from the lungs into the bloodstream, they can cause further internal problems such as cardiovascular disease. These damaging effects are amplified during the *in utero* period. The reduced oxygen or organ damage sustained by the pregnant woman leads to less oxygen transferred to

[7] In fact, to focus regulatory activities on smaller particles, the EPA replaced the earlier TSP air quality standard with a standard for particulate matter less than 10 micrometers in size (PM-10) in 1987. In 1997, the EPA also added a standard for particulate matter less than 2.5 micrometers in size (PM-2.5).

the fetus and impairs fetal brain development. Further, the particulates can be transferred to the fetus directly through the bloodstream and harm the development of fetal respiratory and cardiovascular systems.[8]

All of these *in utero* physiological impacts may translate into damages to cognitive function as a child develops and enters adulthood. Thus, early-life exposure to air pollution may impact long-run human capital formation and adult earnings through both neurological channels as well as direct health impairments (Graff Zivin and Neidell, 2013). For instance, respiratory conditions may have long-term consequences for school attendance, occupational choice, and labor force participation more generally.

Of course, the influence of air pollution on human health and development is not limited to the early-life period. For example, prior research has documented contemporaneous impacts of environmental toxins on adult mortality (Chay *et al.*, 2003) and student test scores (Lavy *et al.*, 2012). The empirical challenge is thus to isolate the long-term effect of air pollution exposure in early childhood from any contemporaneous impacts throughout the lifecycle. A research design that compares individuals born in areas with cleaner air to individuals born in areas with more pollution would not succeed in uniquely identifying the effects of early exposure, since people living in the "treatment" regions may in principle be exposed to lower air pollution over the full life cycle. As highlighted by Heckman and Mosso (2014), finding a long-run effect of early-life exposure to air pollution using this design would be consistent with two possible explanations: (1) a strong initial effect that is attenuated at later stages in the lifecycle, and (2) a weaker initial effect that is amplified at later stages in the lifecycle.

To distinguish between these two channels, one must use a research design that can compare individuals who have different exposure to air pollution in the early-life period, but *the same exposure at older ages*. To formalize this idea, we present a simple framework.[9] Let an individual's health stock be a function of inputs during two time periods: $h = h(I_1, I_2)$, where I_t are inputs in each period t, and $t = 1, 2$. In our case, we can think of $t = 1$ as representing early childhood, while $t = 2$ as representing the rest of life up to the point of observation.

An individual's earnings are a function of his health stock h and his education level e, where education also depends on the health stock. Formally, $y = y(e, h) = y(e(h(I_1, I_2)), h(I_1, I_2))$, where y represents earnings and e represents years of schooling. We are interested in the impact of a change to health inputs in period 1 (I_1) on earnings. We leverage variation from the implementation of CAAA to identify this effect. More precisely, the CAAA lowered air pollution levels in certain counties.

[8]For more information on the physiological pathways by which particulates can impact human health, please see a recent EPA report available at: http://cfpub.epa.gov/ncea/cfm/recordisplay.cfm?deid=216546 (U.S. EPA, 2009).

[9]Our framework is closely related to the model described in Bleakley (2010), which draws upon the framework laid out by Cunha and Heckman (2007). Our model, however, abstracts away from modeling parental investments in response to health shocks (Becker and Tomes, 1976), or the dynamic complementarities between shocks and investments across different time periods (Cunha and Heckman, 2007). We instead focus on the reduced-form relationship between early-life inputs into health and adult earnings because this is what we can measure in our data. The framework could also model pollution as a more general direct input into earnings without hypothesizing that the mechanism occurs solely through the health stock.

Our analysis compares outcomes of cohorts born just before and just after the CAAA in the affected counties (relative to the difference in outcomes between the same cohorts in unaffected counties). In this setting, the treatment group of cohorts born right after the CAAA has "high-quality" inputs (i.e., lower pollution levels) in both periods 1 and 2. By contrast, the comparison group of cohorts born just before the CAAA in affected counties has "low-quality" inputs in period 1 (i.e., high pollution levels), but also experiences the same "high-quality" inputs in period 2 (assuming they continue to live in their counties of birth). Thus, by comparing these two groups, our analysis isolates the *additional* impact of changes to I_1 investments on adult earnings:

$$\frac{\partial y}{\partial I_1} = \left[\frac{\partial y}{\partial e} \times \frac{\partial e}{\partial h} \times \frac{\partial h}{\partial I_1} + \frac{\partial y}{\partial h} \times \frac{\partial h}{\partial I_1} \right] \tag{1}$$

In sum, a change in period 1 investments, I_1, affects health stock, h. This shock to the health stock, in turn, affects long-run earnings through two channels: a direct effect of health on earnings ($\frac{\partial y}{\partial h} \times \frac{\partial h}{\partial I_1}$), and an indirect effect mediated by changes to education ($\frac{\partial y}{\partial e} \times \frac{\partial e}{\partial h} \times \frac{\partial h}{\partial I_1}$).

The goal of the rest of the paper is to deliver estimates of $\frac{\partial y}{\partial I_1}$, where the change in period 1 inputs stems from changes in the levels of ambient air pollution experienced by cohorts surrounding the 1970 CAAA. We also analyze mechanisms that help distinguish between direct effects of health and indirect effects mediated by education. The precise details of both the research design and econometric strategy are described more fully below.

3 The Clean Air Act

The Clean Air Act regulates air pollution in the United States and is the largest environmental program in the country. The Clean Air Act requires the Environmental Protection Agency (EPA) to develop and enforce regulations to protect the general public from exposure to airborne contaminants that are known to be hazardous to human health. The Act was passed in 1963 and significantly amended in 1970, 1977, and 1990. The enactment of the Clean Air Act Amendments of 1970, by authorizing federal regulations to limit emissions, resulted in a major shift in the federal government's role in air pollution control. In doing so, the EPA established national ambient air quality standards (NAAQS), which specify the minimum level of air quality acceptable for six criteria air pollutants.[10]

In a series of path-breaking papers, Henderson (1996) first showed how nonattainment designations lead to large changes in ambient air concentrations; Chay and Greenstone (2003a, 2005) then used these regulatory-induced changes as a source of quasi-experimental variation to better understand the relationships between ambient air pollution, infant health, and willingness to pay for air quality more generally. Particularly relevant to this paper, Chay and Greenstone (2003a) documented that the 1970 CAAA led to a large reduction in ambient air pollution in newly regulated counties, and then showed

[10]These pollutants consist of sulfur dioxide (SO_2), particulates (TSP, PM2.5, and PM10), nitrogen dioxide (NO_2), carbon monoxide (CO), ozone, and lead.

how this reduction significantly lowered infant mortality in the affected counties. We ask whether these same changes in air pollution in the 1970s have any long-run consequences for the cohorts who have survived.

While the Chay and Greenstone (2003a, 2005) papers serve as the underlying basis for our research design, they also presage potential sources of bias in isolating the relationship between early-life air pollution exposure and earnings in adulthood. For example, the CAAA-induced reduction in infant mortality suggests that the wage distribution of surviving cohorts depends on the earnings potential of the "marginal" births that were saved by the air quality improvements. Additionally, there is evidence that CAAA led to increases in housing prices in affected communities (Chay and Greenstone, 2005). This finding at least raises the possibility that some households may have responded to the CAAA by differentially moving in or out of counties with cleaner air. These types of endogenous mobility responses may make cohorts born before and after the changes in air quality in affected counties less comparable. Our research design and robustness tests intend to address these and many other sources of confounding variation. The exact details are specified in subsequent sections.

4 Data

Our primary analysis file combines administrative data from the U.S. Census Bureau's Longitudinal Employer Household Dynamics File (LEHD) with ambient air pollution monitoring data from the EPA. This section describes the datasets, and additional detail can be found in Appendix C.

Air Pollution and Nonattainment Designation

We measure air pollution using data from the EPA's air pollution monitoring network, which provides annual readings for the universe of air pollution monitors scattered throughout the United States. Following Chay and Greenstone (2003a, 2005), we construct two measures of county-level TSP emissions in each year. The first measure is a weighted average of annual TSP emissions over all monitors in a county, with weights proportional to the number of monitor observations within a given year. The second measure is the second highest TSP reading in a county-year. We only use data from monitors that had more than 15 readings in a given year.

These two pollution measures form the basis for the national ambient air quality standards, central to the Clean Air Act and county nonattainment designations. Specifically, the NAAQS designate a county as nonattainment if one of the following criteria are met in a given year: (i) the annual geometric mean concentration exceeds 75 $\mu g/m^3$, or (ii) the second-highest daily concentration exceeds 260 $\mu g/m^3$. As highlighted in Chay and Greenstone (2003a, 2005), the EPA does not maintain historical records of actual county-level nonattainment status dating back prior to 1978. Thus, we classify counties into nonattainment status by applying the NAAQS criteria to their 1970 TSP emissions. We also test the sensitivity of our results to alternative imputations of nonattainment status in Appendix B.

Since not all counties have air pollution monitors, we are able to classify a total of 560 counties—291 nonattainment and 269 attainment—based on their 1970 readings. As we describe below, our main analysis sample uses data from 24 states, and thus we are left with 148 counties—97 nonattainment and 51 attainment—in these states. In most of our specifications, we use data on TSP emissions in these 148 counties over 1969-1974.

Longitudinal Employer Household Dynamics File (LEHD)

The Census Bureau's LEHD file provides administrative quarterly earnings records for over 90% of the United States workforce.[11] The earnings records correspond to the report of an individual's UI-covered earnings by an employer in a given quarter. While the LEHD earnings records are fairly comprehensive, notable exceptions include the self-employed, agricultural workers, and some state, local, and federal employees.

The LEHD provides longitudinal employment and earnings histories for workers along with some basic demographic characteristics such as sex and race.[12] Crucially for our analysis, the LEHD also provides information on both the place and exact date of birth. The place of birth variable in the LEHD is a string variable detailing in most cases the city and state of birth (e.g., "Los Angeles, California"). We developed a matching algorithm to connect this string variable to the Census Bureau database of places, counties, and minor civil divisions as well as the United States Geological Survey's Geographic Names Information System (GNIS) file. We have thus created a crosswalk between the LEHD place of birth string variable and County FIPS codes. A full description of the matching algorithm is detailed in Appendix C.1. Over 95 percent of the individuals in the LEHD file were matched to their county of birth. Lastly, we use the Bureau of Economic Analysis "county-equivalent" as our baseline definition of a county, both to maintain a consistent definition of counties throughout our sample frame as well as to match the BEA's Regional Economic Information System (REIS) data described below.

LEHD Sample Construction. While the LEHD provides extraordinary levels of detail for a large fraction of the United States workforce, it has some important limitations. The LEHD is assembled by combining various states' administrative earnings records. As a result, states have varying degrees of temporal coverage in the main dataset, with most states entering the sample by the late 1990s. Since we only observe earnings records for individuals working in a given year and a given state, we cannot distinguish between non-employment and employment in a state outside the LEHD sample. Put differently, an important caveat to almost any analysis using administrative earnings data is that it is impossible to distinguish between two types of individuals: 1) those who earn zero earnings in any given year because they become unemployed or exit the labor force, and 2) those who

[11] See Abowd *et al.* (2008) and McKinney and Vilhuber (2008) for a comprehensive discussion of both the construction and contents of the LEHD files.

[12] The race variable is divided into 6 mutually exclusive categories: White, Black, Other, Asian, Hispanic, and American Indian.

earn zero earnings simply because they move outside of the states covered in the LEHD. We attempt to construct a sample that reduces the degree to which this issue is relevant. We limit our sample to the 24 states which continuously contain earnings records during 1998-2007, and we limit the sample to individuals who were born in one of those 24 states. Workers are able to move from their state of birth to other states, but they will only be in our sample if they ever work in one of these 24 states from 1998-2007.

If the treatment variable (i.e., early-life exposure to clean air due to the CAAA) is correlated with out-of-state mobility, then any impacts on the extensive margin of earnings may in fact be driven by endogenous mobility rather than employment/labor force participation. While we cannot directly test for endogenous mobility from our 24 states into all other states, we do test for differential mobility responses in two separate ways. First, we examine whether individuals born into nonattainment counties after the CAAA are more or less likely to work in a state other than their birth state. Second, we test whether individuals born into nonattainment counties after the CAAA are more or less likely to move to one of the 6 remaining LEHD states that do not fit our sample restrictions. In both cases, we find no evidence in favor of differential mobility. We also present results omitting zero earnings observations. Intensive margin earnings impacts are subject to less sample attrition bias (see e.g., Jacobson *et al.* (1993); Von Wachter *et al.* (2009); Walker (2013)), and we show that these effects in our analysis, while attenuated, are similar to the main results.

In sum, our baseline sample consists of earnings records for individuals who were born in one of our 24 sample states and who ever worked in one of our 24 states between 1998 and 2007. In most specifications, we limit the sample to individuals born between 1969 and 1974 (three years before and after CAAA implementation), in one of the 148 counties with EPA data continuously defined over this time period. Our final sample size is 5.7 million individuals, which we use to construct a balanced panel of birth-county×birth-year cohort data.

Outcome Variables. As the main outcomes in our analysis, we study mean earnings and the mean number of quarters employed between the ages of 29 and 31. We focus on these ages since the correlation between annual earnings and lifetime income rises rapidly as individuals enter the labor market and begins to stabilize only in the late 20s (this is called "overtaking age" in the literature) (Mincer, 1974; Murphy and Welch, 1990).

To calculate the average annual number of quarters employed and the average annual earnings of an individual between the ages of 29 and 31, we use the following procedure: For each individual in our sample, we calculate the years when he turns 29, 30, and 31, and we search for his earnings record in the employment history file. We take the combined earnings for a worker in a given year, adding over both employers and states (in the event of multiple job spells within a year). We also calculate the number of quarters the worker has positive earnings in a given year (i.e., $\in [0, 4]$). If the earnings record is missing for a particular age category (i.e., because the worker is unemployed or has attritted from the data), we estimate specifications where we either keep this earnings record as missing or

we replace it with a zero. We assign each individual his state of work using the state he is observed working for the first time between age 29 and 31. For the individuals for whom we do not observe earnings records in any quarter between age 29 and 31, we assign the state of work using the state in which he has the most quarterly earnings observations either in future or previous years. In the event that a worker has the same earnings in more than one state, we randomly assign the worker to one of these states.[13]

We study labor market outcomes averaged for each individual over a set of ages rather than outcomes measured at a particular age (e.g., age 30) in order to (i) minimize the residual variance in the observed employment and earnings distributions, and (ii) ameliorate concerns that any effects we see are driven by a contemporaneous economic shock in one particular earnings year. While we would like to analyze labor market outcomes over a larger set of ages throughout the life cycle, we are limited by our data, which is only available for years 1998-2007. Our oldest post-CAAA cohorts are 2007-1972 = 35 years old, while our youngest pre-CAAA cohorts are 1998-1971 = 27 years old. In additional specifications we examine age-specific heterogeneity and summary labor market outcomes averaged over all available ages between 27 and 35.

We express all monetary variables in 2008 dollars, adjusting for inflation using the Consumer Price Index. For each cohort, we cap earnings at age 28 equivalent $100,000 allowing for 2% annual growth in earnings in order to limit the influence of outliers.[14] Mean earnings between the ages of 29 and 31 are $23,563 for individuals born in 1969 (in 2008 dollars).

Additional Outcome Measures. The LEHD contains information on educational attainment based on a statistical match between the 1990 Decennial Census and state earnings records. In addition, the LEHD provides linkages to the Current Population Survey (CPS) for a small subsample of individuals who were interviewed in the 1987-1997 waves of the March CPS. We use the LEHD education measure as well as CPS survey responses to analyze mechanisms behind our main results in Section 7.

Additional County×Year Data

We match the LEHD earnings records to the Regional Economic Information System data from the Bureau of Economic Analysis at the "county-equivalent" by birth-year level. These data contain county-year information on local demographics, economic output, and government expenditures. We use data on population counts, employment, per-capita income, and variables on transfer receipt (i.e. total unemployment spending and total transfer receipts in a birth-county×birth-year).

[13]Since we are primarily focused on the effects of pollution exposure in one's place of birth, the state of work at age 30 is immaterial and a possibly endogenous outcome. We only use information on the state of work in some specifications where we include work-state×earnings-year controls.

[14]We cap earnings in order to remove the influence of outliers. However, our results are not sensitive to winsorizing at other points in the distribution or using unadjusted earnings, as we present later. Specifically, we cap earnings at $100,000 for 28 year olds, $102,000 for 29 year olds, $104,040 for 30 year olds, $106,121 for 31 year olds, and $108,243 for 32 year olds.

We also match all of our data to data from the universe of individual-level natality and mortality files from the National Center from Health Statistics (NCHS). These data provide a rich source of time-varying information on maternal, paternal, and child characteristics for each birth county and birth year. Moreover, these data allow us to examine how infant health responds to adverse environmental conditions for our particular subsample of states and compare our results to those found in Chay and Greenstone (2003a).

Lastly, we bring in data on temperature and precipitation in the county and year of birth from Schlenker and Roberts (2009) to control for any relationships between air pollution and weather. Further details about the data may be found in Appendix C.

5 Econometric Specification

Baseline Econometric Model

Our goal is to estimate the relationship between ambient TSP exposure in early childhood and labor market outcomes measured between ages 29 and 31. Our baseline model has the following form:

$$y_{ct}^{a} = \beta_0 + \beta_1 TSP_{ct} + X_{ct}'\tau + \gamma_c + \eta_{st} + \mu_{ct} \tag{2}$$

where outcome y_{ct}^{a} is either annual earnings or quarters employed for individuals of age a who were born in county c and in year t. TSP_{ct} is the average air pollution concentration in birth county c and year t, weighted by the number of monitor observations in that county×year and measured in μg/m^3. X_{ct} is a vector of time-varying socio-economic, demographic, and climate characteristics in the county and year of birth that may also influence earnings determination and labor force participation at ages 29-31. The exact controls that we use vary across specifications and are described in more detail below. γ_c are county fixed effects that control for time-invariant, unobserved determinants of labor market outcomes for workers born in a particular county, while η_{st} are birth-state×birth-year fixed effects that control for time-varying determinants of long-run outcomes that are common to all individuals born in a particular state×year. The key coefficient of interest, β_1, estimates the effect of a one-unit increase in TSP emissions in a cohort's county in their year of birth on the cohort's average labor market outcomes measured 29 to 31 years later.

Equation (2) is a cohort-based model, which can be estimated using data collapsed to the birth-county×birth-year level. However, there are many micro-level determinants of labor market outcomes that are effectively ignored when collapsing the data to the birth-county×birth-year. We control for some of the observed individual earnings heterogeneity while also aggregating the data to the birth-county×birth-year using a "composition-adjusted" earnings measure.[15] We construct the birth-county×birth-year averages using an auxiliary regression, in which we regress our labor market out-

[15]Similar composition-adjusted, aggregate earnings estimators may be found in Angrist and Lavy (2009); Baker and Fortin (2001); Currie *et al.* (2015); Albouy (2009a,c); Notowidigdo (2011); Shapiro (2006), among others.

comes on individual-level covariates (race, sex, and month of birth) as well as birth-county×birth-year fixed effects. The birth-county×birth-year fixed effects from this regression yield the conditional mean labor market outcomes in a birth-county×birth-year cohort, after controlling for the micro-covariates. We use these conditional means as dependent variables in the cohort model from equation (2). Donald and Lang (2007), among others, show the asymptotic equivalence between this two-step group-means estimator and the micro-data counterpart. In Section 6, we show that the results from our baseline specifications using this composition-adjusted aggregated approach are nearly identical to those produced using the underlying individual-level micro data. The virtue of the composition-adjusted aggregated approach is that it substantially reduces the computational burden of running many regressions with almost 6 million observations, while still controlling for micro-level heterogeneity. In addition, from the stand point of statistical inference, this method allows us to estimate models collapsed to the level of variation, ensuring that the tests are of correct size given serial correlation in the within group errors. Appendix C.2 provides additional details on creating the composition-adjusted birth-county×birth-year mean outcomes.

We estimate equation (2) using weighted least squares, where the weights are the number of individuals in each birth-county×birth-year cell.[16] In all of our regression models, we cluster the standard errors at the commuting zone level to account for any spatial dependence in nonattainment designations within the same metropolitan area.[17] The key coefficient of interest, β_1, now represents the effect of a one-unit increase in TSP emissions in an individual's county and year of birth, holding constant race, sex, and month of birth.

Exposure to TSP in the year of birth is likely correlated with many observable and unobservable determinants of long-run labor market potential. Including birth-county fixed effects γ_c will absorb any time-invariant determinants of long-run human capital unique to a specific county, and including birth-state×birth-year fixed effects will control for transitory determinants of long-run outcomes common to all cohorts born in a given state×year. However, there may exist local *and* transitory determinants of long-run outcomes that also covary with ambient air pollution. For example, local economic conditions are both strong predictors of ambient TSP (Chay and Greenstone, 2003b) and have been shown to affect infant health and fertility decisions (Dehejia and Lleras-Muney, 2004; Lindo, 2011; Schaller, 2012) as well as long-run mortality (Van den Berg *et al.*, 2011; Van Den Berg *et al.*, 2006). Any unobserved transitory local shocks that covary with both TSP and long-run outcomes will lead to bias in the OLS estimate of β_1.

[16] The asymptotic equivalence of our group-level estimator relative to the micro-data regression holds when the group-level weights are the inverse sampling variance of the group coefficients. In practice, it is computationally difficult to recover the sampling variances of the group-level estimates. Thus, we follow Albouy (2009b,c); Angrist and Lavy (2009); Currie *et al.* (2015) (and others) and weight by the group-level cell size (i.e., birth-county×birth-year cell size in our case). Since the sampling variance is proportional to the cell size, we believe this is a reasonable approximation.

[17] The USDA Economic Research Service used county-level commuting data from the 1990 Census data to create 741 clusters of counties that are characterized by strong commuting ties within CZs and weak commuting ties across CZs (Tolbert and Sizer, 1996). Subsequent researchers have used similar levels of Census geography for economic research on local labor markets (e.g. Autor and Dorn (Forthcoming); Walker (2013)).

Using the 1970 CAAA in an Instrumental Variables Design

In order to address concerns about the endogeneity of pollution exposure, we instrument for changes in air pollution using the introduction of the CAAA. Prior research has shown that nonattainment designation is a strong predictor of changes to county-level ambient air pollution (see e.g. Auffhammer *et al.* (2009); Chay and Greenstone (2003a, 2005); Grainger (2012); Henderson (1996); Sanders and Stoecker (2015)). We model this change in air pollution using an indicator 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:

$$TSP_{ct} = \alpha_0 + \alpha_1(Non_{1970,c} \times 1[\tau > 1971]) + X'_{ct}\rho + \gamma_c + \eta_{st} + \nu_{ct} \tag{3}$$

where TSP in a county c and year t is regressed on a time-invariant county indicator equal to 1 if a county is designated as nonattainment, $Non_{1970,c}$, interacted with an indicator equal to 1 in the years after the CAAA went into effect, $1[\tau > 1971]$. This interaction term is equal to 1 for nonattainment counties in the years after CAAA implementation. The other controls are the same as in our baseline OLS model (2). The parameter of interest is α_1, which provides a difference-in-differences estimate of the impact of nonattainment designation on county TSP levels in the years after CAAA regulations went into place.

In the second stage, we use the predicted TSP levels from equation (3) in place of the actual TSP levels in equation (2):

$$y^a_{ct} = \sigma_0 + \sigma_1\widehat{TSP_{ct}} + X'_{ct}\kappa + \gamma_c + \eta_{st} + \varepsilon_{ct} \tag{4}$$

The coefficient of interest, σ_1, represents the effect of a one-unit, CAAA-driven increase in TSP emissions in a cohort's birth-county×birth-year on the cohort's composition-adjusted labor market outcomes measured at ages 29-31.

In subsequent sections, we present evidence that the first stage relationship is strong—nonattainment designation led to significant and persistent declines in ambient TSP concentrations in the years after the law went into effect. We also present evidence that our instrument may satisfy the excludability condition required for a consistent estimate of σ_1. While the identifying assumption is inherently untestable, we conduct several indirect tests. First, we use data from years prior to CAAA implementation to examine pre-trends in county-level covariates and outcomes, finding little evidence of statistically significant differences between nonattainment and attainment counties. We also test whether nonattainment designation is correlated with changes in the observable characteristics of mothers giving birth in the years after the CAAA went into effect, and we find little evidence in support of such hypothesis. Nevertheless, the exclusion restriction may not be perfectly satisfied, as it is plausible that CAAA enactment affected counties in other ways beyond pollution reduction. For instance, prior literature has shown that while nonattainment designation reduces pollution, it does so at the cost of some economic competitiveness (Greenstone, 2002; Greenstone *et al.*, 2012; Walker,

2011, 2013). Therefore, the CAAA may have contributed to declining economic conditions in nonattainment counties, leading to adverse impacts on the long-run earnings capacity of children born into these counties.[18] Such impacts of nonattainment designation on the local economy may lead to bias when interpreting our 2SLS estimates. As a result, we present both the reduced form effects of nonattainment and the IV dose-response estimates throughout. We interpret the reduced form estimates as measuring the overall effects of the CAAA on cohorts born into nonattainment counties in the years after the policy went into effect.

Transitional Dynamics and Distributed Lag Specifications. Equations (3) and (4) implicitly assume that the CAAA improved air quality instantly, and these improvements lasted forever. These models also do not test for heterogeneity in the long-run effects by age of exposure. We estimate distributed lag models to better understand this heterogeneity. We expand our analysis sample to individuals born in years 1969-1977, and we interact indicators for each birth year with the county nonattainment indicator. Thus, the first stage regression model in a distributed lag framework becomes:

$$TSP_{ct} = \zeta_0 + \sum_{k=1969}^{1977} \zeta_k (Non_{1970,c} \times 1[\tau = k]) + X'_{ct}\rho + \gamma_c + \eta_{st} + \omega_{ct} \tag{5}$$

Note that the baseline birth year indicators are absorbed by the birth-state×birth-year fixed effects, η_{st}. The coefficients of interest are the ζ_k's, which estimate the time-path of ambient TSP levels in nonattainment counties before and after the CAAA went into place. In the presence of county fixed effects, not all of the λ_k's are identified, and we make the normalization $\lambda_{1971} = 0$.

We also estimate models that explore similar transitional dynamics for long run labor market outcomes. This results in the following reduced form distributed lag model:

$$y^a_{ct} = \psi_0 + \sum_{k=1969}^{1977} \psi_k (Non_{1970,c} \times 1[\tau = k]) + X'_{ct}\kappa + \gamma_c + \eta_{st} + \varsigma_{ct} \tag{6}$$

The reduced form models also allow us to examine how the CAAA treatment effect varies by age of exposure. Note that all individuals born in 1973 or later in nonattainment counties are exposed to lower TSP from conception onward. Individuals born in 1972 experience lower air pollution in their year of birth—thus, this cohort is partially exposed to cleaner air *in utero* and fully exposed from birth onward. Individuals born in 1971 experience cleaner air from age 1 onward; those born in 1970 experience cleaner air from age 2 onward; and those born in 1969 experience cleaner air from age 3 onward. Since we normalize the coefficient for the 1971 cohort to be zero, our analysis essentially tests for differential effects of exposure relative to exposure at age 1 and older. If there are additional benefits to being exposed to cleaner air between conception and age 1, then we would expect the

[18]The actual impact of nonattainment status on the broader local economy is fairly small. Estimates in Walker (2013) suggests that the implied impact of nonattainment on county employment is less than 0.7% of the total workforce.

coefficients $\psi_{1972} - \psi_{1977}$ to be positive. Similarly, if cleaner air at age 1 has an additional benefit relative to cleaner air at age 2 or older, we would expect coefficients ψ_{1970} and ψ_{1969} to be negative.

6 Results

Cross-Sectional and Fixed Effects Relationships Between TSP Exposure and Long Run Outcomes

We begin by documenting the correlation between pollution in a cohort's county in their year of birth and labor market outcomes measured at ages 29-31. Table 1 presents estimates of equation (2) using various controls. The outcome in Panel A is mean annual quarters of employment $\in [0, 4]$, while the outcome in Panel B is the mean annual earnings. Column (1) presents estimates of the cross-sectional relationship between pollution exposure in the year of birth and later life labor market outcomes, without any controls. The results suggest that individuals born in counties with higher TSP concentrations have lower employment and earnings, although the coefficients are not statistically significant. Purely cross-sectional relationships are likely subject to substantial omitted variable bias—for example, more polluted counties tend to have higher poverty rates, and individuals born in poorer counties have lower earnings capacity in adulthood. These differences are demonstrated in Appendix Table A1, which compares counties with above and below median pollution levels in the years before the policy. Column (3) of Appendix Table A1 presents p-values from a formal test of the difference in means between observable characteristics of these two sets of counties, and we can reject the null hypothesis that the counties are the same for several of the observable variables (at conventional levels of statistical significance).

Columns (2)-(5) of Table 1 include birth-county fixed effects and birth-state×birth-year interactions to control for time-invariant birth-county characteristics as well as any state time-varying factors that may influence both TSP levels and adult outcomes. Columns (2)-(5) also differ in the birth-county time-varying control variables that are included, with the set of controls increasing in stringency as one moves from left to right across the table. Column (2) adds in flexible controls for climate and weather in the birth-county×birth-year to absorb some of the potentially confounding relationships between temperature, precipitation, and ambient air pollution.[19] Columns (3) and (4) include further controls for time-varying birth, maternal, and paternal characteristics in a birth-county×birth-year from birth certificates data.[20]

[19] Weather controls include a linear, quadratic, and cubic terms in annual county precipitation. We also include flexible temperature controls, calculated as the number of "degree days" in a given county year above 0, 5, 8, 10, 12, 15, 20, 25, 29, 30, 31, 32, 33, 34 degrees Celsius (i.e. 14 separate terms).

[20] Controls in the "Natality Basic" column include a continuous measure of both mother and father education, mother's age, and indicators for marital status of mother, month of the first prenatal care, and an indicator for no prenatal care. Controls for the "Natality Unrestricted" columns include the "Natality Basic controls" in addition to: indicators for years of education of both the mother and father (<12, =12, 13-15, 16+), father's age, indicators for mother's age (10-14, 15-19, 20-24, 25-29, 30-34, 35-39, 40+), delivery outside of hospital indicator, physician present at delivery indicator, previous live

Unlike the cross-sectional model in column (1), the fixed effects OLS models in columns (2)-(5) point to a *positive* relationship between TSP exposure in the year of birth and long-run labor market outcomes. These models control for all time-invariant characteristics of counties that may predict later life well-being, and the identifying variation comes from within-county changes in pollution levels. Appendix Table A2 shows that observable 1969 characteristics of counties that had above and below median changes in TSP between 1970 and 1972 are similar along several margins. We no longer see as many statistically significant differences in baseline characteristics as we did in Appendix Table A1. However, some important differences remain—for instance, counties with below-median changes in TSP have higher average levels of parental education relative to counties with above-median changes in TSP. More generally, as noted above, fixed effects regressions cannot control for all time-varying forms of endogeneity. For example, areas with increases in pollution may also be experiencing upward trends in economic activity, which may have independent influences on the human capital attainment of cohorts born in these areas. We next turn to using the CAAA as a source of quasi-experimental variation for identifying the long-run causal effects of early-life exposure to air pollution.

Using the CAAA in a Quasi-Experimental Design

We begin by presenting evidence of the first stage relationship between CAAA implementation and air pollution levels in nonattainment counties. The results in Appendix Table A3 correspond to estimates of α_1 in equation (3). Consistent with the previous literature, we find a strong relationship between CAAA implementation and ambient concentrations of TSP in nonattainment counties. This relationship is robust across specifications and suggests that CAAA reduced TSP concentrations by 8-12 $\mu g/m^3$. Relative to a mean value of 95.9 $\mu g/m^3$, this amounts to about a 10 percent reduction in air pollution for the average county in our sample.

Figure 1 plots the coefficients from the distributed lag model in equation (5). This regression specification mimics the controls in column (5) of Appendix Table A3, although the sample frame is widened to span the years of 1969-1977. Consistent with the results in Appendix Table A3, we see a persistent decline in ambient TSP in the years after the CAAA went into effect.

We can also use the distributed lag model to examine trends between treatment and control counties in the years prior to policy implementation. We see that in the years immediately preceding CAAA initiation, trends between eventual nonattainment and attainment counties evolve similarly.[21] The figure provides suggestive evidence that changes in attainment counties serve as a useful counterfactual for what would have happened to nonattainment counties in the absence of the regulation, a key condition for identification in a difference-in-difference estimator.

Table 2 provides additional evidence for the validity of our research design. Columns (1) and (2) of

birth indicator (1, 2+), previous fetal death indicator (1, 2+), last pregnancy was live birth indicator, last pregnancy was fetal death indicator, indicators for 1-11, 12-17, 18 or more months since last live birth, indicators for 1-11, 12-17, 18 or more months since termination of last pregnancy.

[21] The first year for which we have data on air pollution is 1969, and we are thus unable to examine pre-trends before 1969.

Table 2 present means of observable characteristics for both attainment and nonattainment counties in 1969, whereas Columns (3) and (4) present the same statistics in log differences between 1969 and 1971. Columns (5) and (6) present p-values from tests of the null hypotheses that the levels and pre-trends in characteristics of attainment and nonattainment counties are not statistically different. While Column (5) makes clear that nonattainment counties are observably different than attainment counties, Column (6) suggests that *trends* in observable characteristics between attainment and nonattainment counties are similar in the years prior to the 1970 CAAA. Across most specifications, we cannot reject the null hypothesis that the difference in trends is zero. These results suggest that cohorts in attainment counties may serve as valid counterfactuals for cohorts born in nonattainment counties. There is one covariate which exhibits significant differences—total transfers per capita. Columns (3) and (4) suggest that nonattainment counties exhibit about 2% more growth in pre-period total per capita transfers. As a result, we attempt to control flexibly for total transfers per capita in all regressions by interacting the pre-determined county per capita transfers from 1969 with quadratic polynomial trends.

Table 3 presents estimates from regressions that modify equation (2) by replacing the independent variable of interest, TSP_{ct}, with county nonattainment status interacted with an indicator for a post-1971 birth year, $(Non_{1970,c} \times 1[\tau > 1971])$. These reduced form regression models show how CAAA implementation affected labor market outcomes of individuals born into nonattainment counties 29-31 years later. Panel A presents results using mean annual quarters of employment in a birth-county×birth-year cell as the dependent variable. As in Table 1, we add in more controls as we move across the columns. Column (1), which includes birth-county fixed effects and birth-state×birth-year interactions but no other controls, shows that cohorts born into nonattainment counties in the years after CAAA went into effect work on average 0.020 quarters more, relative to the counterfactual. Relative to a mean number of employed quarters of 2.74, this effect amounts to a 0.7 percent increase in quarters employed. The coefficients are very similar when we include controls for the weather (column (2)), and for time-varying characteristics from birth certificates data (columns (3) and (4)).

Panel B of Table 3 presents results using mean annual earnings in a birth-county×birth-year as the dependent variable. Relative to baseline earnings in nonattainment counties of $23,623, our estimates suggest that CAAA implementation increased the earnings of cohorts born into "cleaner" counties by about 1 percent.[22] The estimates in Panel B are slightly larger in magnitude than those in Panel A; an increase in labor force participation by 0.020 quarters is equivalent to only about $117 in annual earnings, suggesting that the estimated impact on earnings is driven by both extensive and intensive margin effects.[23] In other words, the effect on annual earnings may be driven by both an increase in the number of quarters worked as well as higher productivity while in the labor force or higher attachment to the labor force that does not show up in quarters worked (e.g., going from part-time

[22]The average effect across all columns in Panel B of Table 3 is $259, where the average is taken across columns, weighting by the inverse of the standard error.

[23]This statistic is calculated by noting that 1 quarter is equal to 91.25 days out of the year. The average daily earnings is $64 ($23,623/365). Hence, $64×91.25×0.020=$116.8.

employment to full employment, working in every quarter).

Panels C and D of Table 3 directly estimate the intensive margin, or non-zero, earnings effect by limiting the sample to workers with strictly positive earnings between ages 29 and 31. Panel C shows the effect on mean annual log earnings, and Panel D shows the effect on mean annual non-zero earnings in levels. Since the results above suggest that most of the estimated impacts occur along the labor force participation margin, we expect these coefficients to be smaller in magnitude than the coefficients in Panel B, but to remain positive. The log earnings specifications in Panel C show that CAAA implementation is associated with a 1 percent increase in annual earnings. When we study non-zero earnings in levels rather than logs, we still find positive coefficients, although they are smaller in magnitude and not statistically significant. The difference between the log and level specifications suggests that the log-linear function may better fit the distribution of non-zero earnings. In sum, Panels C and D suggest that the earnings effects we found in Panels A and B are also present for workers with non-zero earnings, although a majority of the effect is occurring along the extensive work/non-work margin.

Figure 2 presents graphs based on estimates of distributed lag models described in equation (6), using quarters worked and annual earnings as dependent variables, respectively. There are two main findings: First, we see that trends leading up to the CAAA between treatment and control cohorts are nearly identical for both outcomes. We view this as additional evidence in support of the identifying assumption in the model—that trends in outcomes between treatment and control groups would have evolved similarly except through the change in policy. Second, in the years after the CAAA we see a mean shift in both quarters worked and annual earnings.[24] The transitional dynamics in the figure point to additional benefits of exposure to cleaner air between conception and age 1, relative to exposure at age 1 and later. By contrast, the lack of negative coefficients for cohorts born in 1969 and 1970 suggests that there are little to no differential benefits from exposure beginning at ages 1, 2, or 3. Appendix Figure A1 presents the distributed lag analysis for the intensive margin earnings measures. As presaged by the results in Table 3, comparisons between Figure 2 and Appendix Figure A1 suggest that most but not all of the impact is occurring along the extensive work margin; non-zero earnings and log earnings exhibit weaker but still positive mean shifts in the years after the CAAA.[25]

Table 4 presents IV estimates of equation (4). The results suggest that a ten-unit increase in ambient TSP exposure in the year of birth reduces average annual age 29-31 earnings by around 1 percent. The IV estimates in Table 4 are equal in magnitude (but of opposite sign) to the reduced form estimates in Table 3. This result is to be expected since the first stage model in Appendix Table A3 showed a decrease in ambient TSP of around 10 $\mu g/m^3$, and the TSP variable in Panel B is scaled by a factor of 10, representing a 10 $\mu g/m^3$ increase in ambient TSP. As before, the estimates improve in statistical precision as we reduce the residual variance in long-run earnings determination

[24] P-values for tests of the difference in means between the coefficients before and after the CAAA in each graph are equal to 0.013 and 0.045, respectively.

[25] P-values for tests of the difference in means between the coefficients before and after the CAAA in each graph are equal to 0.195 and 0.081, respectively.

by including additional control variables.

As mentioned above, all of our models are estimated using composition-adjusted labor market outcomes, aggregated to the birth-county×birth-year cohort level. We have also estimated the baseline regressions using the individual micro-level data. The results, presented in Appendix Tables A4 and A5, yield nearly identical coefficients to those presented in our main results, with slightly smaller standard errors.

Treatment Effect Heterogeneity

We explore treatment effect heterogeneity in a variety of ways. Table 5 presents estimates using annual earnings and quarters worked measured at ages 27 through 35 as outcomes. Each column corresponds to a different regression using a different age-earnings sample. As we move across the columns, the cohorts we can include in our sample change—e.g., we can only observe cohorts born in 1971-1974 at age 27 (column 1), and we can only observe cohorts born in 1969-1972 at age 35 (column 9). We can observe our baseline sample of cohorts born in 1969-1974 at ages 29-33 in our data (columns 3-7). Column (10) presents estimates from a summary index earnings measure taken over all years 27-35. At each age of observation that we can measure, the results are qualitatively consistent with the baseline results from before; being born in a nonattainment county in the years after CAAA improves long-run labor market outcomes. Although the coefficients are not identical across age categories, the confidence intervals overlap across all of them. This pattern is also seen in Figure 3, where we present the coefficients and 95% confidence intervals from these regressions graphically. The similarity in coefficients likely reflects the fact that earnings and employment are highly correlated across ages. These results also demonstrate that (i) positive effects on labor market outcomes are found at more than one (ultimately, somewhat arbitrary) age category, and (ii) our results are not confounded by a contemporaneous change in earnings determinants in later years. As evidence of the latter point, consider that Columns (1)-(9) are estimated using the same individuals whose earnings are recorded in different years (e.g., cohorts born in 1971 show up between 1998 (Column 1) and 2006 (Column 9)). The earnings measure in Column (10) serves as a type of "summary-index" of labor market outcomes over a nine-year age span and reduces the residual variance in annual earnings. The one downside with the 27-35 earnings index relative to our baseline 29-31 earnings index is the sample imbalance that occurs in early and later ages; for ages less than 28 we lose pre-CAAA time periods, and for ages older than 33 we lose post-CAAA time periods. Table 6 presents the corresponding IV estimates for the same set of outcomes, and the results are consistent with both the results in Tables 4 and 5; higher TSP in the year of birth is associated with lower long-run earnings capacity.[26]

Next, we examine heterogeneity in effects across the earnings distribution. We estimate a series

[26]We have also estimated specifications where we use labor market outcomes averaged over sets of ages as outcomes, but we control for age fixed effects in the first-step auxiliary regression before aggregation. In other words, we take out fixed effects for single years of age before taking the average of the residualized outcomes over a set of ages. We have estimated these models for workers aged 29-31 and workers aged 27-35. The results are nearly identical to the baseline estimates and are available in Appendix B.

of regression models that explore how CAAA implementation and TSP exposure affect the fractions of cohorts in various percentiles of the earnings distribution.[27] We begin by calculating the 1st, 5th, 10th, 25th, 50th, 75th, 90th, 95th, and 99th percentiles of the within-county earnings distribution for the 1969 birth cohort. For each subsequent cohort born in 1970-1974, we classify individuals into bins based on their place in the "pre-treatment" 1969 within-county earnings distribution. We calculate the fraction of individuals from a given birth-county×birth-year cohort who are in each bin of the pre-treatment earnings distribution (e.g., the fraction of individuals whose earnings place them below the 1st percentile of their county's 1969 distribution; the fraction of individuals whose earnings place them between the 1st and 5th percentiles of their county's 1969 distribution; etc.). Table 7 presents results of regression models that use the fraction of workers in each quantile of the 1969 cohort earnings distribution as a dependent variable. The results are graphically summarized in Figure 4. The estimates suggest that most of the mean earnings effect is driven by the bottom tail of the distribution; CAAA implementation is associated with a relative decrease in the fraction of individuals with earnings at the bottom tail of the distribution and a relative increase in the fractions in middle parts of the distribution. These results suggest that changes in the extensive/participation margin explain most of the observed earnings impacts. This finding is also consistent with prior literature showing how *in-utero* shocks lead to increased disability rates for adults 60 years later (e.g., Almond, 2006), which may translate into weaker labor force attachment. We further explore the mechanisms underlying these effects in Section 7.

Finally, we explore heterogeneity by race and gender by interacting our key treatment variable with indicators for different race and gender categories using the underlying micro data. Since these models contain three-way interaction terms (Race/Sex×($Non_{1970,c} \times 1[\tau > 1971]$)), we also include the additional lower-order interaction terms (i.e., Race/Sex×$Non_{1970,c}$ and Race/Sex×$1[\tau > 1971]$)). Tables 8 and 9 present results for earnings and quarters worked, respectively. In each table, only one of the interaction coefficients is marginally significant at the 10% level, suggesting little to no treatment effect heterogeneity across race and gender groups.

Composition, Selection, and Alternative Research Designs

We explore the robustness of our results to a variety of additional tests and specifications. We highlight some of the main results here and relegate additional analysis and discussion to Appendix B.

Testing for Changes in Population Characteristics

An important concern for our study is that improvements in air quality might change the composition of the population in nonattainment counties, leading to changes in the characteristics of the children born in them. For example, families may respond to the CAAA by differentially moving in or out of the counties with clean air. This is particularly relevant as Chay and Greenstone (2005) find that

[27] Sample sizes preclude us from estimating quantile treatment effects directly using our micro data.

nonattainment designation is associated with increases in housing values nearly 10 years after the legislation went into effect. If these increases in housing values reflect that higher socio-economic status families are migrating to counties with cleaner air (Banzhaf and Walsh, 2008), then we may observe changes in the underlying population characteristics of nonattainment counties post-CAAA. This would imply that the positive impacts on long-run earnings capacity may be in part driven by changes in the types of individuals giving birth in nonattainment counties rather than a causal effect of early-life exposure to cleaner air.

Table 10 investigates whether CAAA led to a compositional shift in the underlying population in nonattainment counties. Each column represents a different dependent variable. Columns (1)-(3) use data from the NCHS Vital Statistics records to estimate whether the maternal education or the fraction of white or black children differentially change in the years after nonattainment designation.[28] Column (4) uses data from the BEA to estimate whether nonattainment status is correlated with differential changes in per-capita income in newly regulated counties. Lastly, Column (5) uses the LEHD earnings records to form a predictive earnings index based on sex and race of workers.[29] The results in Table 10 provide little evidence for differential sorting along observables that might bias our estimates. The point estimates are not only statistically insignificant, but also small in magnitude, and the signs of the coefficients suggest our estimates, if anything, may be slightly downwardly biased.

Although CAAA implementation did not lead to changes in observable population characteristics, there may be sorting along unobserved margins. However, as it takes time to move, we might expect that most migration responses would only materialize in a few years after the CAAA. To reduce the likelihood that unobservable compositional changes may be biasing our results, we explore the sensitivity of our main estimates to restricting the sample to cohorts born in 1970-1972, the years immediately surrounding policy implementation. Appendix Table A6 presents the reduced form results of the effect of nonattainment designation in the year of birth on adult labor market outcomes, and the results are similar to our baseline estimates and remain statistically significant.[30] Appendix Table A7 presents results from the IV models for the same 1970-1972 window. The coefficients are somewhat larger in magnitude than in our baseline models, but they are less well estimated. The first stage F-statistics, pertaining to the strength of the instrument at predicting variation in air pollution in this shortened time window, are all below 10. Thus, the imprecision of the 2SLS estimates may also reflect attenuation associated with weak instruments (Staiger and Stock, 1997; Stock *et al.*, 2002).

We also explore whether nonattainment designation affects fertility or the total number of workers observed in the data in Appendix Table A8. Columns (1) and (2) present results using log(# births)

[28]Maternal education was not reported by all states during our analysis time frame. In our data over 1969-1974, 9 out of the 24 sample states did not report maternal education. Counties in these states are omitted from the analysis of maternal education.

[29]Specifically, we use the micro data to estimate earnings regressions controlling for sex and race indicators. We then use the predicted values from this regression as a summary index measure of sorting in Column (5).

[30]For the 1970-1972 specifications, we replace our 1st-4th order baseline polynomial trends with linear trends in pre-determined (1969) employment, population, total transfers per capita, and unemployment transfers per capita. The full set of quartic interactions cannot be fit with 3 years of data.

and log(# workers) in a birth-county×birth-year as dependent variables (unweighted). Columns (3) and (4) use the sex ratio at birth and the ratio of male to female workers in the data as outcomes, respectively.[31] We find some evidence that more workers in the LEHD are born in nonattainment counties in the years after the 1970 CAAA, although the estimated effect is only marginally significant In column (5), we examine the ratio of the total number of workers in a birth-county×birth-year in the LEHD over the total number of births in that county×year, finding similar results to Column (1); the ratio of workers to births increases in nonattainment counties in the years after the CAAA. Note that the observed increase in workers is consistent with the main findings of increased labor force participation.

Lastly, being born into a nonattainment county after CAAA may lead to differential migration patterns across states. This issue is especially relevant for our analysis as we only observe individuals who ever appear in one of the 24 sample states from our baseline sample. Consequently, our results may be biased if the CAAA affects the likelihood that an individual moves out of the set of sample states between when he appears in our data (i.e. has positive earnings) and the time of observation we use for our measure of labor market outcomes (ages 29-31). We address this concern in two ways. First, we examine the relationship between CAAA implementation and mobility into the 6 other states in the LEHD not in our baseline sample. Second, we examine the relationship between CAAA implementation and out-of-birth-state-mobility within our 24 sample states. Panel A of Appendix Table A9 presents results where the dependent variable is the fraction of individuals in a cohort who work in one of the six LEHD states not in our baseline sample. Panel B of Appendix Table A9 presents results where the dependent variable is the fraction of individuals in a cohort who are working at ages 29-31 in a state in the LEHD other than the state in which they born. Both panels suggest that the relationship between mobility and CAAA implementation is unlikely to be a significant source of bias, with confidence intervals ruling out even a small amount of differential mobility.

Results from a Regression Discontinuity Design

The results we have presented are based on a difference-in-difference design stemming from changes in TSP levels across nonattainment and attainment counties. However, since CAAA regulations apply non-linearly in the county TSP level, it is possible to exploit this non-linearity for identification using a regression discontinuity design (RDD) (Chay and Greenstone, 2005). The RDD thought experiment focuses on cohorts born in counties with 1970 TSP levels that were just above the 75 μg/m^3 nonattainment threshold and cohorts born in counties with 1970 TSP levels just below.[32] Since we are interested in the additional effects of exposure to clean air in very early childhood (relative

[31] Population sex ratios may be impacted by CAAA if there are effects on fetal deaths (see Sanders and Stoecker, 2015 for evidence on this topic).

[32] As noted above, areas were designated as nonattainment if they violated either of two conditions in their TSP readings: (1) the annual geometric mean was great than 75μg/m^3, or (2) the second highest reading for the year was greater than 260μg/m^3. In practice, however, the annual geometric mean standard was binding and the second highest reading standard was not binding. Only 2 out of our 97 nonattainment counties satisfied (2) and not (1).

to exposure at slightly older ages), we implement the RDD analysis by comparing the difference in outcomes between cohorts born in counties just below and above the nonattainment threshold in the years after CAAA went into effect to the difference in outcomes between cohorts born in the same counties in the years before the CAAA.[33]

The validity of an RDD rests on two additional assumptions: (i) counties cannot precisely manipulate their pre-CAAA pollution levels to fall below the nonattainment threshold (no sorting), and (ii) all other county characteristics are smooth functions of the running variable at the threshold. We find evidence in support of these assumptions: a formal density test (McCrary, 2008) fails to reject the null hypothesis that the density is smooth across the threshold, and predetermined county characteristics do not discontinuously change at the threshold (see Appendix Table A10).[34]

We estimate the RDD model using our main analysis sample of cohorts born in the 24 sample states between 1969 and 1974. We augment equation (3) with a linear spline in the 1970 annual geometric mean of the TSP level (i.e., we include a linear term in the 1970 TSP level and also the interaction between this variable and county nonattainment status).[35] For cohorts born in years prior to 1972, the running variable is set to zero. Panels A and B of Appendix Table A11 present results from the local linear regressions using counties with 1970 TSP levels in three different bandwidths surrounding the nonattainment threshold: 50 μg/m^3, 100 μg/m^3, and 150 μg/m^3. Columns (1)-(3) present results from the RDD first stage, with county TSP as the dependent variable. Columns (4)-(6) and (7)-(9) present estimates using earnings and quarters worked as outcome variables, respectively. Results from the cross-validation procedure following Lee and Lemieux (2010) indicate a bandwidth that uses the full sample (i.e. the 150 bandwidth) is most appropriate, which is unsurprising given the small sample size in the analysis.

There are two primary conclusions. First, the magnitudes from the RDD estimates are similar in size to our baseline difference-in-differences estimates; we observe negative impacts of nonattainment on TSP levels and positive impacts on long-run labor market outcomes, with the bandwidth suggested by the cross-validation yielding marginally significant estimates. Second, the results are somewhat sensitive to the choice of bandwidth and the choice of polynomial in the running variable.[36] We attribute this sensitivity of the RDD estimator to a lack of density in the running variable; with only 148 counties in the entire sample, and fewer around the threshold, it is difficult for the model to identify polynomial terms on either side of the discontinuity with sufficient precision. Recent work by Sanders and Stoecker (2015) reaches similar conclusions with respect to the 1970 CAAA RDD

[33] RDD analyses in panel settings of this nature have been used in the prior literature (e.g., Cellini *et al.*, 2010; Isen, 2014.)

[34] Due to Census disclosure rules, we are unable to disclose RDD figures and therefore only present results from regression analyses.

[35] Lower order polynomials in running variables have been shown to perform better in RD settings relative to higher order polynomials (Gelman and Imbens, 2014).

[36] We have also experimented with higher order polynomials. The specification that minimizes the Akaike Information Criterion (AIC) and Bayes Information Criterion (BIC) is a fifth order polynomial, which yields a \$365 and 0.017 impact on earnings and quarters worked, respectively, albeit both are above conventional significance levels ($p \approx 0.13$). However, the coefficient on TSP is even less significant.

design. We therefore view these results as complementary to our main evidence, but we believe the difference-in-difference model is preferred, given the lack of density on either side of the threshold.

Mobility and Differential Pollution Exposure over the Life Cycle

We have interpreted our estimates as capturing the effects of additional exposure to clean air in one's year of birth relative to exposure at slightly older ages. However, this interpretation is subject to an implicit assumption that our treatment is uncorrelated with subsequent mobility patterns. If, for example, cohorts exposed to cleaner air from age 0 onwards were more likely than comparison cohorts to subsequently move to other counties with even cleaner air, then our main estimates would capture the combined effects of early-life and later-life exposure to cleaner air. While we have already shown that treated cohorts are no more or less likely to move away from their state of birth to one of the other states in our sample in Appendix Table A9, we explore whether later-life pollution exposure is mediating our findings in two additional ways.

First, we examine whether individuals born into nonattainment counties in the years after CAAA implementation have differential exposure to pollution in adulthood. Since we only observe individuals' states of work (and not their states or counties of residence in adulthood), we measure later-life pollution exposure using the population-weighted average ambient pollution level in each person's state of work at ages 29-31. Column (1) of Appendix Table A12 uses age 29-31 TSP exposure as an outcome, while column (2) uses age 29-31 PM-10 exposure as an outcome.[37] We find no correlation between being born into a nonattainment county in the years after the CAAA and either type of pollution exposure in adulthood. Although the measures of adult pollution exposure are imperfect (albeit the best we can construct with our data), these results provide suggestive evidence that differential pollution exposure later in life is unlikely to be mediating our main results.

Second, we examine if there are differential treatment effects depending on whether an individual works in his state of birth or whether he migrates to another state. In Appendix Table A13, we estimate our baseline reduced form and IV regressions separately for individuals who are working in their birth state at age 30 and for individuals who are not working in their birth state at age 30. Columns (1) and (2) explore treatment effect heterogeneity for earnings, whereas Columns (3) and (4) explore treatment effect heterogeneity for quarters worked. Panel A presents estimates from the reduced form, whereas Panel B presents 2SLS estimates. We find no statistically significant heterogeneity across these groups—the 95 percent confidence intervals overlap for both "stayers" and "leavers" for each outcome. In sum, we conclude that our main estimates are unlikely to be affected by differential mobility responses and/or differential later life pollution exposure.

[37] During the years when we observe our cohorts in adulthood (1998-2007), the EPA transitioned from monitoring TSP to focusing on smaller particulates such as PM-10. As a result, the number of TSP monitors declines over the sample period, and the number of PM-10 monitors rises, and we present results using both measures as dependent variables.

Infant Mortality and Non-Random Selection in Later Life Earnings

In an influential paper, Chay and Greenstone (2003a) showed how CAAA-induced changes in TSP lowered infant mortality rates in affected counties. As a result, cohorts born before CAAA may have different outcomes from cohorts born after, not because of the relationship between TSP and human capital formation, but instead due to selection: the marginal infants "saved" by cleaner air may have different earnings potential that could pull the cohort earnings average either up or down. To assess the degree to which our estimates might suffer from this type of bias, we first replicate Chay and Greenstone (2003a) for our subset of counties and states. Appendix Tables A14 and A15 present results for both the 1969-1974 window as well as the 1970-1972 window, respectively. The estimates from Appendix Table A14 suggest that CAAA implementation is associated with a decrease in mean birth weight and a reduction in the infant mortality rate.[38] However, the latter effect is not statistically significant at conventional levels. These two results are not inconsistent with one another, as a reduction in the infant mortality rate typically "saves" marginal children that are more likely to have low birth weight. Appendix Table A15 shows that regression models focusing on a narrower window before and after the 1970 CAAA show a statistically significant decrease in the infant mortality rate. The results in Appendix Table A15 suggest the 1970 CAAA are associated with a 13.8 percent decrease in the infant mortality rate in newly regulated counties in our sample.[39] The average nonattainment county in our sample had 102 infant deaths in 1972, which suggests that roughly 5 additional lives were saved per county. Relative to a mean county cohort size of around 2990, this suggests the mortality-induced selection effect should be relatively small (i.e., an increase in the size of the labor force of 0.16 percent).

Nevertheless, we implement a bounding exercise following Lee (2009). The idea is to estimate our main regressions using a sample of individuals who would have survived regardless of CAAA implementation. We trim our analysis sample by the number of "extra" individuals who are selected into the sample (i.e., who do not die) as a result of the CAAA.[40] The results are presented in Appendix Table A16, where Columns (1) and (2) present results for earnings, and Columns (3) and (4) present results for quarters worked. Columns (1) and (3) present estimates pertaining to the lower bound, whereas Columns (2) and (4) present estimates pertaining to the upper bound. The lower bound estimates are quite close to our preferred estimates, although the slight attenuation in the coefficients with similar standard errors renders the earnings result no longer statistically significant, although the quarters worked result remains so. The upper bound estimates are also contained within the confidence intervals of our preferred specifications, which suggests that the mortality-induced selection effect is

[38] The internal infant mortality rate is defined as the number of infant deaths due to internal causes divided by the number of live births in a given year.

[39] Chay and Greenstone (2003a) find about a 3-6 percent decrease in the infant mortality rate for the full set of nonattainment counties.

[40] To calculate the lower bound of the effect on each outcome, we drop the top 0.16 percent of individuals in the distribution of the post-CAAA, nonattainment cohorts. To calculate the upper bound of the effect on each outcome, we drop the bottom 0.16 percent of individuals in the distribution of each post-CAAA, nonattainment cohorts.

likely quite small.

7 Interpretation and Discussion

Our analysis thus far suggests that exposure to cleaner air *in utero* and before age 1 improves an individual's adult labor market performance measured some 30 years later. However, up to this point, we have provided very little direct evidence on the mechanisms by which these effects might be operating. We extend our analysis by presenting some suggestive analyses exploring mechanisms in three additional ways. First, we examine the effects of CAAA nonattainment designation on our measure of educational attainment from the LEHD. Second, we turn to survey data from the 2000 Census (5% sample) and the 2005-2012 American Communities Survey (ACS) available through the Integrated Public-Use Microdata Series (IPUMS) to examine poverty status and welfare benefit receipt in adulthood. Lastly, we rely on a heretofore underutilized linkage between the LEHD and the March CPS to explore additional later life outcomes, including disability.

Although the LEHD provides relatively little information on individual characteristics (primarily, sex and race), it does contain information on educational attainment that is based on a statistical match to the 1990 Decennial Census.[41] We augment equation (2) by replacing the dependent variable with the LEHD continuous measure of years of schooling. Appendix Table A17 presents results, which suggest that cohorts born into nonattainment counties in the years after CAAA have about 0.02 more years of education relative to the control group. This is a small effect size, especially when compared with our estimated impact on long-run earnings. If the return to an additional year of schooling is about 10 percent (Card, 1999), and the effects of CAAA on long-run earnings occurred solely through increases in educational attainment, we would only expect a $0.02 \times 10 = 0.2$ percent increase in earnings. By contrast, we find a 1 percent increase in age 29-31 earnings, implying that the benefits of cleaner air *in utero* and in early life may also operate through channels that affect adult outcomes beyond education, such as health and non-cognitive ability.

In order to explore these mechanisms further, we turn to the 2000 Census and the ACS, which provides survey responses on welfare income receipt and poverty status.[42] We estimate regressions similar to our baseline analyses, focusing on individuals born in 1969-1974 in the 24 LEHD sample states, weighting by population.[43] One important difference between these models and the models presented before is that the Census and ACS do not contain information on individual county of

[41] See Vilhuber and McKinney (2014) for additional information on the LEHD education variable.

[42] The 2000 Census and ACS also include information on educational attainment. The effects on educational attainment in the Census and ACS are consistent with the findings in Appendix Table A17, although not as precisely estimated.

[43] We implement the same 2-step estimation method, where the first step auxiliary regression controls for indicators for sex, race, survey year, and the full set of birth-county×birth-year dummies. We then use the coefficient estimates of the birth-county×birth-year dummy variables as the dependent variables, estimating reduced form versions of equation (4), controlling for birth-county fixed effects, birth-state×birth-year fixed effects, predetermined county characteristics (population, employment, total transfers) interacted with quadratic trends, and the other county time-varying controls described above. We weight these regressions by the sum of the person weights in each birth-county×birth-year cell (the results are similar when we weight by the number of individuals in each cell), and cluster standard errors on the commuting zone level.

birth or date of birth. Therefore, we use an individual's county of residence at the time of survey as his or her county of birth, following Sanders (2012).[44] The challenge is that county of residence is different than county of birth for many individuals given the high rates of geographic mobility in the United States. Thus, assigning early childhood exposure based on later life locations will introduce measurement error into the underlying treatment variable. In addition, we must also impute year of birth using respondents' ages at the time of the survey. Appendix Table A18 presents results. There is suggestive evidence that treated cohorts are less likely to live below the poverty line in adulthood, consistent with our finding that labor market impacts are concentrated in the bottom of the earnings distribution. However, this evidence comes with the caveats associated with using the Census and ACS mentioned above.

Lastly, we use a linkage between the LEHD and the CPS for a small subsample of individuals who were surveyed in the 1987-1997 waves of the March CPS. Our cohorts, which were born in 1969-1974, are observed between age 13 and 28 in these data. We use the linked LEHD-CPS subsample to explore three additional outcomes: self-reported annual earnings, the self-reported amount of income received from unemployment insurance, and the self-reported amount of income received from disability insurance. Appendix Table A19 presents the results from reduced form models estimating the effects of CAAA implementation in the year of birth on cohorts born in nonattainment counties. The coefficients are generally not very precise, possibly due to the much smaller sample sizes in this analysis. We do however find a marginally significant decline in disability income receipt, which suggests that at least part of our estimated impact on long-run earnings may operate through health channels.[45] We also still see a positive (albeit insignificant) coefficient on self-reported earnings, which is consistent with our main results from the administrative records in the LEHD.

In sum, our analysis of mechanisms across three data sets provides suggestive evidence that the CAAA-induced reductions in air pollution likely improved both cognitive ability as well as health capital. We have also shown that the earnings effects are mostly (but not entirely) driven by the extensive margin of labor force participation, and the increases in earnings are exhibited by individuals moving from the bottom to the median of the earnings distribution (see e.g., Figure 4). The CPS points to one reason why this may be happening—the cohorts born in nonattainment counties after the CAAA are less likely to receive disability insurance income relative to the cohorts born before. In addition, Bharadwaj *et al.* (2014) and Sanders (2012) provide suggestive evidence that cleaner air in early childhood improves elementary school and high school test scores, respectively. It is therefore reasonable to expect that improvements in both health and cognitive ability translate into higher earnings potential in adulthood.

[44] Not all counties are identifiable in the Census and ACS data. Only large counties that are coterminous with lower levels of geography such as state economic areas (SEAs), county groups, or public-use microdata areas (PUMAs) are identifiable. See https://usa.ipums.org/usa-action/variables/COUNTY for more details. We are able to use 121 out of the 148 counties in the LEHD states with complete EPA data for this analysis.

[45] When correcting p-values for multiple hypothesis testing across these outcomes, column (3) is no longer statistically significant.

The Long Run Benefits of the 1970 Clean Air Act Amendments

In this paper, we have asked whether early-life exposure to air pollution impacts long-term labor market outcomes. The evidence suggests the answer is "Yes," but there are some important caveats. The thought experiment has been to compare individuals born in counties that experienced large regulatory-induced changes in ambient TSP exposure, before and after the regulations went into place. As suggested by Figure 1, the declines in TSP were persistent in the years after CAAA implementation. This means that individuals born before CAAA went into effect (i.e., in years 1969-1971) likely experienced some of the benefits of reduced TSP exposure beginning at ages 1 (for those born in 1971), 2 (for those born in 1970), and 3 (for those born in 1969). However, these individuals were not afforded the luxury of exposure to cleaner air between conception and age 1 as did those born in 1973-1974. Thus, as highlighted above, our estimates point to *additional* benefits of exposure between conception and age 1, and cannot speak to any cumulative effects of early childhood exposure to pollution (e.g., between ages 0 and 5). While our distributed lag models suggest that exposure at age 1 does not have any *differential* impacts relative to exposure at age 2 or 3, they do not imply that there are no benefits at all. If there are in fact further benefits of exposure at slightly older ages in early childhood, then we are underestimating this cumulative impact.

There are also other sources of bias that may cause us to underestimate the effect of early-life air pollution exposure (i.e., downwardly bias our results as opposed to biasing our results towards zero). For example, nonattainment designation influences production decisions of firms, resulting in negative impacts on county-level economic activity (Greenstone, 2002; Walker, 2011, 2013). There is some evidence that economic conditions during gestation affect both fertility (Dehejia and Lleras-Muney, 2004) and infant health outcomes (Lindo, 2011; Johnson and Schoeni, 2011). However, the available estimates from Walker (2011, 2013) suggest that the effect of nonattainment designation on the overall labor market is small.[46] Additionally, a regression of contemporaneous county-level per capita income on CAAA enactment, as presented in Table 10, failed to detect an effect, which also suggests at most a minor effect on the overall local economy.

Our baseline regression estimates suggest that nonattainment designation (or a 10 μg/m^3 reduction in TSP more generally) leads to a 1 percent increase in earnings at age 29-31, or $260 annually. This magnitude represents about one tenth of the estimated return to an extra year of schooling in the U.S. (Card, 1999). It is also at the lower end of the existing range of estimates of the long-run returns to early-life shocks: For example, Almond (2006) finds that *in-utero* exposure to the 1918 flu epidemic is correlated with a 5-9 percent reduction in men's age 40-60 earnings, Chetty *et al.* (2011) show that a high-experience kindergarten teacher is associated with a 7 percent increase in age-27 earnings, while Heckman *et al.* (2010) estimate a 15 percent increase in lifetime earnings resulting from the Perry Preschool intervention. Our estimate may be smaller because we study a less severe shock to the early-life environment and due to the downward biases discussed above.

[46] An important caveat is that Walker (2013) studies the 1990 CAAA, which may have had more modest labor market effects than the original 1970 CAAA.

To interpret the magnitude of the effect of air pollution in the year of birth on earnings at ages 29-31, we calculate the lifetime earnings impact implied by our estimates. We assume that the percentage gain in earnings remains constant at 1 percent over the life cycle and that earnings are discounted at a 3 percent real rate (i.e., a 5 percent discount rate with 2 percent wage growth) back to age zero.[47] Under these assumptions, the mean present value of lifetime earnings at age zero in the U.S. population is approximately $434,000. We calculate this number using the mean wage earnings from the March 2008 Current Population Survey to obtain an earnings profile over the lifecycle.[48] Thus, the financial value of being born into a nonattainment county in the years after CAAA went into effect is 1 percent of $434,000 or $4,340 per person.

In 1972, there were approximately 1.5 million births in newly designated nonattainment counties, implying that the total increase in lifetime earnings for this cohort amounted to about $6.5 billion (2008 dollars) annually. To the extent that this reduction in TSPs was permanent (recall Figure 1), these benefits would have accrued in each of the 40 years since then. In this case and assuming a 5 percent discount rate, the present discounted value of these earnings gains is around $118 billion (2008 dollars).[49] Moreover, these calculations ignore any of the non-wage amenity benefits associated with cleaner air as well as the benefits of any further reductions in TSPs caused by the Clean Air Act in later years. Of course, a full cost-benefit analysis also requires precise information about the costs of these regulations.[50]

8 Conclusion

In this paper, we provide the first empirical quasi-experimental examination of the relationship between individuals' *in-utero* and early childhood exposure to environmental toxins and their labor market outcomes measured 30 years later. We exploit variation induced by the introduction of the CAAA in 1970, which imposed restrictions on the maximum-allowable concentrations of TSP emissions in nonattainment counties. Our analysis compares cohorts in nonattainment counties born just before and after the legislation-mandated reductions in air pollution relative to the same difference among cohorts in attainment counties.

We find that an individual's exposure to lower ambient air pollution levels in the year of birth positively impacts earnings 30 years later. Specifically, we show that the approximate ten percent reduction in TSP that resulted from CAAA implementation is associated with a one percent increase in age-30 earnings among affected cohorts in our sample states. Assuming a constant effect over the lifecycle, we calculate an approximate $4300 average cumulative lifetime income gain in present value

[47] This type of forecasting was originally pioneered by Theil (1958).

[48] Based on this exercise, the cumulative (non-discounted) average lifetime earnings is around $1.5 million (2008$). Poterba *et al.* (2010) estimate that average lifetime earnings in the United States are $1.6 million (2010$).

[49] This calculation assumes that otherwise, pollution levels as well as the marginal effect of pollution would have remained the same for the last forty years (e.g., technology and supply-demand factors also remained the same).

[50] See Greenstone *et al.* (2012) and Walker (2013) for recent evidence on the costs associated with the Clean Air Act.

terms, implying that early-life air quality contributes a total of $6.5 billion in lifetime earnings for each affected cohort. We present suggestive evidence that these labor market effects manifest through both improvements in educational attainment (perhaps through improvements in cognitive ability) and improvements in later life health measures (as identified through reductions in disability receipts).

Disadvantaged populations disproportionately live in more polluted areas in the United States. This fact has complicated research examining the effects of pollution on health, as disadvantaged individuals, who also have worse health outcomes, tend to be exposed to higher levels of air pollution because of their residential location. However, these persistent disparities also raise interesting questions: why does the pollution-health relationship remain, and how does this persistence influence broader social and economic inequality in this country? Our analysis suggests that both economic and environmental inequality may be reinforcing. If this circular relationship is robust, then policies designed to improve air quality may also play a role in thwarting this cycle, thus serving not only as environmental health policies but also as effective social policies for reducing economic disparities.

Our paper additionally contributes to a large and growing literature in economics that documents a lasting relationship between early-life conditions and adult well-being by studying a policy-driven shock to the pre- and post-natal environment of modern U.S. cohorts. Previous work on this topic has been challenged by data limitations, with very few large-scale datasets that combine detailed information on location and date of birth together with adult outcomes. We introduce the LEHD data set as a new resource for studying these issues in the United States.

References

Abowd, J. M., Stephens, B. E., Vilhuber, L., Andersson, F., McKinney, K. L., Roemer, M. and Woodcock, S. (2008). The lehd infrastructure files and the creation of the quarterly workforce indicators. *LEHD Working Paper*.

Adhvaryu, A. and Nyshadham, A. (2011). *Endowments and investments within the household: Evidence from iodine supplementation in Tanzania*. Discussion Paper 998, Economic Growth Center.

Aizer, A. and Cunha, F. (2010). Child endowments, parental investments and the development of human capital, Brown University, unpublished manuscript.

Albouy, D. (2009a). The unequal geographic burden of federal taxation. *Journal of Political Economy*, **117** (4), 635–667.

— (2009b). The unequal geographic burden of federal taxation. *Journal of Political Economy*, **117** (4), 635–667.

— (2009c). *What are cities worth? Land rents, local productivity, and the capitalization of amenity values*. Tech. rep., National Bureau of Economic Research.

Almond, D. (2006). Is the 1918 influenza pandemic over? long-term effects of in utero influenza exposure in the post-1940 us population. *Journal of Political Economy*, **114** (4), 672–712.

— and Currie, J. (2011). Human capital development before age five. In O. Ashenfleter and D. Card (eds.), *Handbook of Labor Economics*, vol. 4, Elsevier, pp. 1315–1486.

—, Edlund, L., Li, H. and Zhang, J. (2010). Long-term effects of the 1959-1961 china famine: Mainland china and hong kong. In *In The Economic Consequences of Demographic Change in East Asia*, University of Chicago Press, pp. 321–350.

—, — and Palme, M. (2009). Chernobyl's subclinical legacy: Prenatal exposure to radioactive fallout and school outcomes in sweden. *The Quarterly Journal of Economics*, **124** (4), 1729–1772.

— and Mazumder, B. (2012). *Fetal origins and Parental Responses*. Working Paper 2012-14, Federal Reserve Board of Chicago.

Angrist, J. and Lavy, V. (2009). The effects of high stakes high school achievement awards: Evidence from a randomized trial. *American Economic Review*, **99** (4), 1384–1414.

Auffhammer, M., Bento, A. and Lowe, S. (2009). Measuring the effects of the clean air act amendments on ambient pm10 concentrations: The critical importance of a spatially disaggregated analysis. *Journal of Environmental Economics and Management*, **58** (1), 15–26.

Autor, D. and Dorn, D. (Forthcoming). The growth of low skill service jobs and the polarization of the u.s. labor market. *American Economic Review*.

Baker, M. and Fortin, N. (2001). Occupational gender composition and wages in canada, 1987–1988. *Canadian Journal of Economics*, **34** (2), 345–376.

Banzhaf, S. H. and Walsh, R. P. (2008). Do people vote with their feet? an empirical test of tiebout's mechanism. *The American Economic Review*, **98** (3), 843–863.

Barker, D. J. (1990). The fetal and infant origins of adult disease. *BMJ: British Medical Journal*, **301** (6761), 1111.

Becker, G. S. and Tomes, N. (1976). Child endowments and the quantity and quality of children. *The Journal of Political Economy*, **84** (4), S143–S162.

Bharadwaj, P., Eberhard, J. and Neilson, C. (2013a). Health at birth, parental investments and academic outcomes, UC San Diego, unpublished manuscript.

—, Gibson, M., Zivin, J. G. and Neilson, C. A. (2014). *Gray Matters: Fetal Pollution Exposure and Human Capital Formation*. Working Paper 20662, National Bureau of Economic Research.

—, Løken, K. V. and Neilson, C. (2013b). Early life health interventions and academic achievement. *American Economic Review*, **103** (5), 1862–91.

Black, S., Devereux, P. and Salvanes, K. (2007). From the cradle to the labor market? the effect of birth weight on adult outcomes. *The Quarterly Journal of Economics*, **122** (1), 409–439.

Bleakley, H. (2010). Health, human capital, and development. *Annual Review of Economics*, **2** (1), 283–310.

Card, D. (1999). The causal effect of education on earnings. *Handbook of labor economics*, **3**, 1801–1863.

Cellini, S. R., Ferreira, F. and Rothstein, J. (2010). The value of school facility investments: Evidence from a dynamic regression discontinuity design. *The Quarterly Journal of Economics*, **125** (1), 215–261.

Chay, K., Dobkin, C. and Greenstone, M. (2003). The clean air act of 1970 and adult mortality. *Journal of Risk and Uncertainty*, **27** (3), 279–300.

— and Greenstone, M. (2003a). *Air quality, infant mortality, and the Clean Air Act of 1970*. Working Paper 10053, National Bureau of Economic Research.

— and — (2003b). The impact of air pollution on infant mortality: Evidence from geographic variation in pollution shocks induced by a recession. *The Quarterly Journal of Economics*, **118** (3), 1121–1167.

— and — (2005). Does air quality matter? evidence from the housing market. *Journal of Political Economy*, **113** (2).

Chetty, R., Friedman, J. N., Hilger, N., Saez, E., Schanzenbach, D. W. and Yagan, D. (2011). How does your kindergarten classroom affect your earnings? evidence from project star. *Quaterly Journal of Economics*, **126** (4), 749–804.

Conti, G., Heckman, J. J., Yi, J. and Zhang, J. (Forthcoming). Early health shocks, parental responses, and child outcomes. *Economic Journal*.

Cunha, F. and Heckman, J. (2007). The technology of skill formation. *The American Economic Review*, **97** (2), 31–47.

Currie, J. (2009). Healthy, wealthy, and wise: Socioeconomic status, poor health in childhood, and human capital development. *Journal of Economic Literature*, pp. 87–122.

—, Davis, L. W., Greenstone, M. and Walker, R. (2015). Do housing prices reflect environmental health risks? evidence from more than 1600 toxic plant openings and closings. *American Economic Review*, **Forthcoming**.

—, Hanushek, E. A., Kahn, E. M., Neidell, M. and Rivkin, S. G. (2009a). Does pollution increase school absences? *The Review of Economics and Statistics*, **91** (4), 682–694.

— and Moretti, E. (2007). Biology as destiny? short-and long-run determinants of intergenerational transmission of birth weight. *Journal of Labor Economics*, **25** (2), 231–264.

— and Neidell, M. (2005). Air pollution and infant health: What can we learn from california's recent experience? *Quaterly Journal of Economics*, (3), 1003–1030.

—, — and Schmieder, J. (2009b). Air pollution and infant health: Lessons from new jersey. *Journal of health economics*, **28** (3), 688–703.

—, Stabile, M., Manivong, P. and Roos, L. L. (2010). Child health and young adult outcomes. *Journal of Human Resources*, **45** (3), 517–548.

— and Walker, R. (2011). Traffic congestion and infant health: Evidence from e-zpass. *American Economic Journal: Applied Economics*, **3** (1), 65–90.

Dahl, G. B. and Lochner, L. (2012). The impact of family income on child achievement: Evidence from the earned income tax credit. *The American Economic Review*, **102** (5), 1927–1956.

Datar, A., Kilburn, M. R. and Loughran, D. S. (2010). Endowments and parental investments in infancy and early childhood. *Demography*, **47** (1), 145–162.

Dehejia, R. and Lleras-Muney, A. (2004). Booms, busts, and babies' health. *The Quarterly Journal of Economics*, **119** (3), 1091–1130.

Del Bono, E., Ermisch, J. and Francesconi, M. (2012). Intrafamily resource allocations: a dynamic structural model of birth weight. *Journal of Labor Economics*, **30** (3), 657–706.

Donald, S. and Lang, K. (2007). Inference with difference-in-differences and other panel data. *Review of Economics and Statistics*, **89** (2), 221–233.

Gelber, A. and Isen, A. (2013). Children's schooling and parents' behavior: Evidence from the head start impact study. *Journal of Public Economics*.

Gelman, A. and Imbens, G. (2014). *Why high-order polynomials should not be used in regression discontinuity designs*. Working Paper 20405, National Bureau of Economic Research.

Gilliland, F. D., Berhane, K., Rappaport, E. B., Thomas, D. C., Avol, E., Gauderman, W. J., London, S. J., Margolis, H. G., McConnell, R., Islam, K. T. *et al.* (2001). The effects of ambient air pollution on school absenteeism due to respiratory illnesses. *Epidemiology*, **12** (1), 43.

Gluckman, P. D. and Hanson, M. A. (2004). Living with the past: evolution, development, and patterns of disease. *Science*, **305** (5691), 1733–1736.

Graff Zivin, J. and Neidell, M. (2013). Environment, health, and human capital. *Journal of Economic Literature*, **51** (3), 689–730.

Graff-Zivin, J. and Neidell, M. (2013). The impact of pollution on worker productivity. *American Economic Review*.

Grainger, C. (2012). The distributional effects of pollution regulations: Do renters fully pay for cleaner air? *Journal of Public Economics*.

Greenstone, M. (2002). The impacts of environmental regulations on industrial activity: Evidence from the 1970 and 1977 clean air act amendments and the census of manufactures. *Journal of Political Economy*, **110** (6).

—, List, J. and Syverson, C. (2012). *The Effects of Environmental Regulation on the Competitiveness of US Manufacturing*. Tech. rep., National Bureau of Economic Research.

Hanna, R. and Oliva, P. (2011). *The Effect of Pollution on Labor Supply: Evidence from a Natural Experiment in Mexico City*. Working Paper 17302, National Bureau of Economic Research.

Heckman, J. J., Moon, S. H., Pinto, R., Savelyev, P. A. and Yavitz, A. (2010). The rate of return to the highscope perry preschool program. *Journal of Public Economics*, **94** (1), 114–128.

— and Mosso, S. (2014). The economics of human development and social mobility. *Annual Review of Economics*, **6**, 689–733.

Henderson, V. (1996). Effects of air quality regulation. *The American Economic Review*, **86** (4), 789–813.

Hoynes, H., Schanzenbach, D. and Almond, D. (2012). *Long Run Impacts of Childhood Access to the Safety Net*. Tech. rep., National Bureau of Economic Research.

Isen, A. (2014). Do local government fiscal spillovers exist? evidence from counties, municipalities, and school districts. *Journal of Public Economics*, **110**, 57–73.

Jacobson, L. S., LaLonde, R. J. and Sullivan, D. G. (1993). Earnings losses of displaced workers. *The American Economic Review*, pp. 685–709.

Johnson, R. (2011). *Long-Run Impacts of School Desegregation and School Quality on Adult Attainments*. Working Paper 16664, National Bureau of Economic Research.

Johnson, R. C. and Schoeni, R. F. (2007). The influence of early-life events on human capital, health status, and labor market outcomes over the life course.

— and — (2011). The influence of early-life events on human capital, health status, and labor market outcomes over the life course. *The BE journal of economic analysis & policy*, **11** (3).

Lavy, V., Ebenstein, A. and Roth, S. (2012). The impact of air pollution on cognitive performance and human capital formation.

Lee, D. (2009). Training, wages, and sample selection: Estimating sharp bounds on treatment effects. *Review of Economic Studies*, **76** (3), 1071–1102.

Lee, D. S. and Lemieux, T. (2010). Regression discontinuity designs in economics. *Journal of Economic Literature*, **48**, 281–355.

Lindo, J. (2011). Parental job loss and infant health. *Journal of Health Economics*, **30** (5), 869–879.

McCrary, J. (2008). Manipulation of the running variable in the regression discontinuity design: A density test. *Journal of Econometrics*, **142** (2), 698–714.

McKinney, K. L. and Vilhuber, L. (2008). Lehd infrastructure files in the census rdc - overview. *LEHD Working Paper*, **Revision : 219**.

Mincer, J. (1974). Schooling, experience, and earnings. *National Bureau of Economic Research*.

Murphy, K. M. and Topel, R. H. (2002). Estimation and inference in two-step econometric models. *Journal of Business & Economic Statistics*, **20** (1), 88–97.

— and Welch, F. (1990). Empirical age-earnings profiles. *Journal of Labor economics*, pp. 202–229.

Nilsson, P. (2009). The long run effects of early childhood lead exposure: Evidence from the phase-out of leaded gasoline. *Working Paper*.

Notowidigdo, M. (2011). *The incidence of local labor demand shocks*. Tech. rep., National Bureau of Economic Research.

Oreopoulos, P., Stabile, M., Walld, R. and Roos, L. L. (2008). Short-, medium-, and long-term consequences of poor infant health an analysis using siblings and twins. *Journal of Human Resources*, **43** (1), 88–138.

Poterba, J., Venti, S. and Wise, D. (2010). The rise of 401 (k) plans, lifetime earnings, and wealth at retirement. *National Bureau of Economic Research*.

Ransom, M. R. and Pope, C. A. (1992). Elementary school absences and pm¡ sub¿ 10¡/sub¿ pollution in utah valley. *Environmental research*, **58** (1), 204–219.

Rau, T., Reyes, L. and Urzua, S. (2013). The long-term effects of early lead exposure: Evidence from a case of environmental negligence, University of Maryland, working paper.

Reyes, J. W. (2007). Environmental policy as social policy? the impact of childhood lead exposure on crime. *The BE Journal of Economic Analysis & Policy*, **7** (1).

Romer, P. M. (1986). Increasing returns and long-run growth. *The Journal of Political Economy*, pp. 1002–1037.

Sanders, N. (2012). What doesn't kill you makes you weaker prenatal pollution exposure and educational outcomes. *Journal of Human Resources*, **47** (3), 826–850.

Sanders, N. J. and Stoecker, C. (2015). Where have all the young men gone? using sex ratios to measure fetal death rates. *Journal of Health Economics*.

Schaller, J. (2012). Booms, busts, and fertility: Testing the becker model using gender-specific labor demand, University of Arizona, unpublished manuscript.

Schlenker, W. and Roberts, M. (2009). Nonlinear temperature effects indicate severe damages to us crop yields under climate change. *Proceedings of the National Academy of Sciences*, **106** (37), 15594–15598.

— and Walker, W. R. (2011). *Airports, air pollution, and contemporaneous health*. Tech. rep., National Bureau of Economic Research.

Schultz, T. W. (1961). Investment in human capital. *The American Economic Review*, pp. 1–17.

Shapiro, J. (2006). Smart cities: Quality of life, productivity, and the growth effects of human capital. *The Review of Economics and Statistics*, **88** (2), 324–335.

Staiger, D. and Stock, J. H. (1997). Instrumental variables regression with weak instruments. *Econometrica*, **65** (3), 557–586.

Stock, J. H., Wright, J. H. and Yogo, M. (2002). A survey of weak instruments and weak identification in generalized method of moments. *Journal of Business & Economic Statistics*, **20** (4).

Theil, H. (1958). *Economic Forecasts and Public Policy*. Amsterdam: North-Holland Publishing Co.

Tolbert, C. M. and Sizer, M. (1996). U.s. commuting zones and labor market areas. a 1990 update. *Economic Research Service Staff Paper*, (No. 9614).

U.S. EPA (2009). *Integrated Science Assessment for Particulate Matter (Final Report)*. Report, U.S. Environmental Protection Agency, Washington, D.C.

Van den Berg, G., Doblhammer-Reiter, G. and Christensen, K. (2011). Being born under adverse economic conditions leads to a higher cardiovascular mortality rate later in life: Evidence based on individuals born at different stages of the business cycle. *Demography*, **48** (2), 507–530.

Van Den Berg, G., Lindeboom, M. and Portrait, F. (2006). Economic conditions early in life and individual mortality. *The American Economic Review*, pp. 290–302.

Vilhuber, L. and McKinney, K. (2014). *LEHD Infrastructure files in the Census RDC-Overview*. Tech. rep.

Von Wachter, T., Song, J. and Manchester, J. (2009). Long-term earnings losses due to mass layoffs during the 1982 recession: An analysis using us administrative data from 1974 to 2004. *Unpublished paper, Columbia University. http://www. columbia. edu/˜ vw2112/papers/mass_layoffs_1982. pdf*.

Walker, R. (2011). Environmental regulation and labor reallocation. *American Economic Review: Papers and Proceedings*, **101** (2).

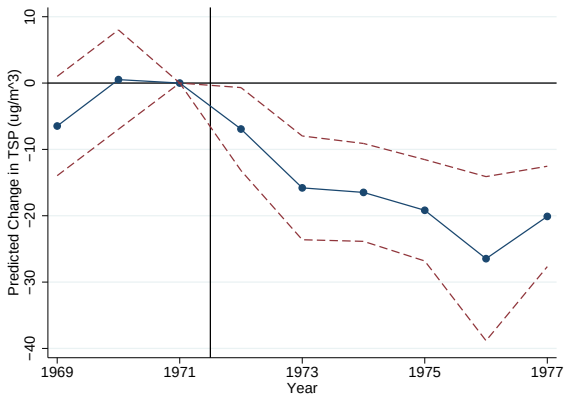
— (2013). The transitional costs of sectoral reallocation: Evidence from the clean air act and the workforce*. *The Quarterly Journal of Economics*, p. qjt022.

Wooldridge, J. M. (1997). On two stage least squares estimation of the average treatment effect in a random coefficient model. *Economics Letters*, **56** (2), 129–133.

— (2002). *Econometric analysis of cross section and panel data*. The MIT press.

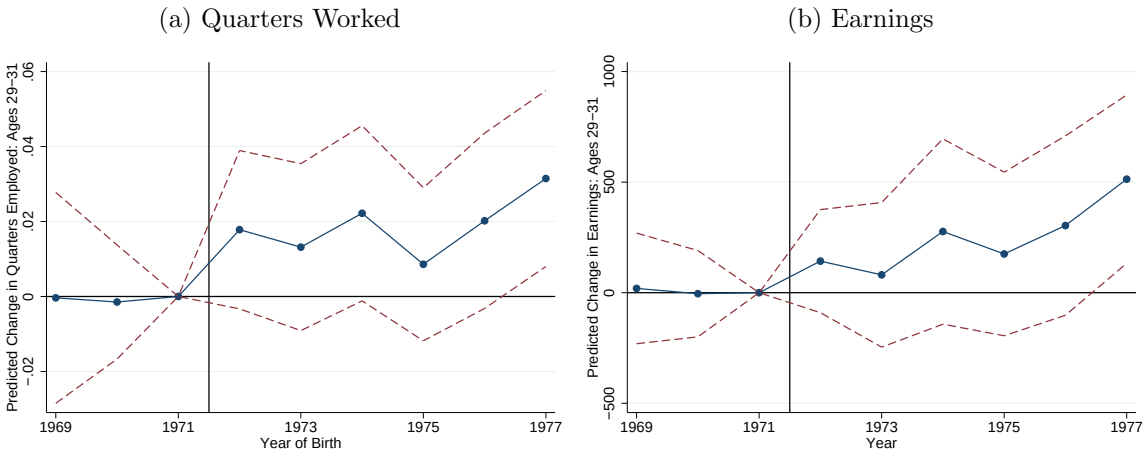
— (2003). Further results on instrumental variables estimation of average treatment effects in the correlated random coefficient model. *Economics Letters*, **79** (2), 185–191.

Figure 1: The Effect of CAAA Implementation on County TSP Levels by Year



NOTE: This figure presents regression coefficients from a version of equation (5). We regress the county-level annual mean TSP on a full set of year indicators interacted with an indicator for county nonattainment status, county fixed effects, state×year fixed effects, county economic controls (population, employment, total transfers), weather controls, and unrestricted natality controls, weighting by the county×year cell size. We plot the coefficients on the interactions between year indicators and nonattainment. The coefficient on the interaction between year 1971 and nonattainment is normalized to zero. The dashed lines represent 95% confidence intervals, where standard errors are clustered by commuting zone. See text for details. Source: EPA.

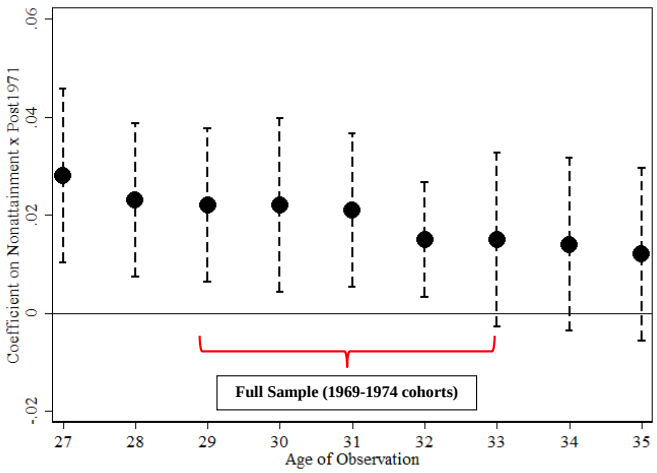
Figure 2: The Effect of CAAA Implementation on Long-Run Labor Market Outcomes at Ages 29-31 by Year of Birth



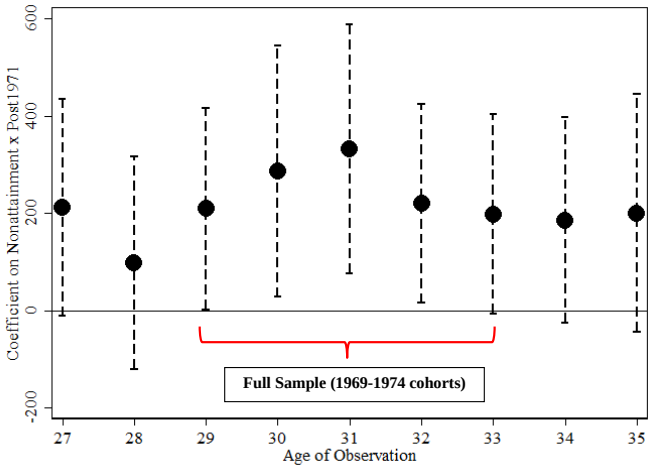
NOTE: This figure presents regression coefficients from a version of equation (6). The dependent variables are indicated in the figure sub-headings. We regress the dependent variable on a full set of birth-year indicators interacted with an indicator for birth-county nonattainment status, birth-county fixed effects, birth-state×birth-year fixed effects, birth-county economic controls (population, employment, total transfers), weather controls, and unrestricted natality controls, weighting by the birth-county×birth-year cell size. We plot the coefficients on the interactions between birth-year indicators and nonattainment. The coefficient on the interaction between birth-year 1971 and nonattainment is normalized to zero. The dashed lines represent 95% confidence intervals, where standard errors are clustered by commuting zone. Source: LEHD, EPA.

Figure 3: The Effect of CAAA Implementation on Long-Run Labor Market Outcomes by Age of Observation

(a) Quarters Worked

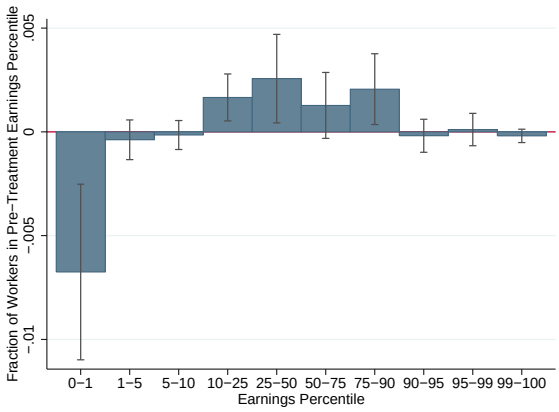


(b) Earnings



NOTE: These figures present regression coefficients and 95% confidence intervals (in dashed lines) from 9 separate regressions for each outcome (annual quarters worked and annual earnings), which are observed at different follow-up ages between age 27 and 35. The plotted regression coefficients are on the interaction between nonattainment county and birth-year 1972 and later. The samples change as different ages of follow-up are considered. See Table 5 for more details. Source: LEHD, EPA.

Figure 4: Heterogeneous Effects of CAAA Implementation Across the Earnings Distribution



NOTE: This figure plots the coefficient estimates from 10 separate regressions, where the dependent variable in each regression is the proportion of individuals in each birth-county×birth-year cohort whose age 29-31 earnings falls within the 1969 earnings quantile indicated along the x-axis. The plotted coefficients correspond to estimates of the reduced form effect of CAAA on the location of individuals within the pre-treatment earnings distribution. All regressions control for birth-county fixed effects, birth-state×birth-year fixed effects, birth-county economic controls (population, employment, total transfers), weather controls, and unrestricted natality controls, weighting by the birth-county×birth-year cell size. The vertical lines represent 95% confidence intervals, where standard errors are clustered by commuting zone. See Table 7 for point estimates. Source: LEHD, EPA.

Table 1: OLS Estimates: The Effect of TSP Exposure in the Year of Birth on Labor Market Outcomes at Ages 29-31

	(1)	(2)	(3)	(4)	(5)
	Panel A: Quarters Employed				
Mean TSPs (/10)	-0.003	0.001	0.001*	0.001***	0.001***
	(0.003)	(0.000)	(0.000)	(0.000)	(0.000)
	Panel B: Annual Earnings				
Mean TSPs (/10)	-100.30	11.76*	12.67*	12.99**	11.88**
	(83.60)	(7.08)	(6.58)	(5.29)	(5.89)
Counties	148	148	148	148	148
First Year	1969	1969	1969	1969	1969
Last Year	1974	1974	1974	1974	1974
Sample Size	888	888	888	888	888
Birth-County FE and Birth-State×Birth-Year FE		Yes	Yes	Yes	Yes
Year of Birth Weather Controls			Yes	Yes	Yes
Year of Birth Natality Basic				Yes	Yes
Year of Birth Natality Unrestricted					Yes

NOTE: This table reports regression coefficients from 10 separate regressions, 5 per panel. The dependent variable in each regression is a composition-adjusted labor market average for a birth-county×birth-year cell. The outcome in Panel A is the mean annual quarters employed $\in [0, 4]$ between ages 29 and 31, and the outcome in Panel B is mean annual earnings between ages 29 and 31. The table reports regression estimates of the effects TSP levels in the year of birth on the outcomes of interest. Column (1) reports does not include any additional controls. In columns (2)-(5), the regressions control for birth-county fixed effects, birth-state×birth-year fixed effects, and predetermined county characteristics (population, employment, total transfers) interacted with quadratic trends. Additional birth-county time-varying controls are listed in the table and described further in the text. The regressions are weighted by the birth-county×birth-year cell size. Standard errors are clustered by commuting zone and are in parentheses. See text for details. Source: LEHD, EPA.

Table 2: Summary Means in 1969 and Pre-Trend Differences Between 1969-1971

	(1) Nonattain	(2) Attainment	(3) Nonattain (1969-1971)	(4) Attainment (1969-1971)	(5) (2)-(1) p-value	(6) (4)-(3) p-value
Cohort Averages: Ages 29-31 (Source: LEHD)						
Earnings (2008$)	23623	23294	-0.024	0.027	0.095	0.834
Earnings: 4-Quarter (2008$)	37109	36654	0.022	0.037	0.107	0.185
Quarters Employed	2.74	2.741	-0.038	-0.017	0.650	0.482
Fraction Working in Non-Birth State	0.31	0.30	-0.007	-0.030	0.879	0.129
County Environment: Age 0 (Source: EPA, Schlenker and Roberts (2009))						
Total Suspended Particulate ($\mu g/m^3$)	95.89	58.59	-0.058	0.044	0.000	0.662
Average Annual Precipitation	988.1	1164	-0.034	-0.041	0.243	0.811
Average Temperature	12.51	12.78	0.020	0.007	0.047	0.485
County Socio-Economic: Age 0 (Source: BEA)						
Income Per Capita (2008$)	17181	16613	0.037	0.035	0.178	0.233
Total Employment	194709	67169	0.025	0.034	0.041	0.265
Transfers UI Per Capita	4.374	4.33	0.243	0.224	0.131	0.304
Total Transfers Per Capita	7.35	7.36	0.031	0.029	0.564	0.002
Cohort Demographics: Age 0 (Source: NCHS)						
% White	0.85	0.86	-0.011	-0.002	0.779	0.903
% African American	0.14	0.11	0.055	-0.002	0.612	0.160
Mother's Education	11.82	11.73	0.001	-0.008	0.588	0.249
Father's Education	12.31	12.2	0.012	-0.016	0.464	0.127
Mother's Age	24.3	24.11	-0.007	-0.009	0.148	0.839
Father's Age	27.45	27.35	-0.011	-0.011	0.321	0.915
Totals						
Workers Born in 1969 working in 1998-2000	749,000	167,000				
Counties	97	51				

NOTE: This is a table of summary means from our baseline estimation sample. The summary statistics are calculated using data from 1969 for a single cross-section of counties. An observation is a county, and the means have been weighted by county population. All dollar amounts are in 2008 dollars. Column (1) presents summary statistics for those counties that exceeded the EPA's NAAQS TSP standard based on monitor readings in 1970. Column (2) presents summary statistics for the subset of counties that were in compliance with the NAAQS standards in 1970. Columns (3) and (4) present log differences between 1969 and 1971 for nonattainment and attainment counties, respectively. Column (5) presents the p-values from a test of the null hypothesis that the levels in columns (1) and (2) are the same in 1969. Column (6) presents the p-values from a test of the null hypothesis that the log differences between 1969 and 1971 are the same for both nonattainment and attainment counties. The p-values in Columns (5) and (6) are generated from a regression of either levels or log differences on an indicator for nonattainment, using robust standard errors for inference, and weighting by the county population.

Table 3: The Effect of CAAA Implementation on Long-Run Labor Market Outcomes at Ages 29-31

	(1)	(2)	(3)	(4)
	Panel A: Quarters Employed			
Nonattainment×1($\tau > 1971$)	0.020**	0.023***	0.019**	0.022***
	(0.010)	(0.008)	(0.008)	(0.008)
	Panel B: Annual Earnings			
Nonattainment×1($\tau > 1971$)	224.59	278.74**	250.57**	276.66**
	(148.36)	(126.23)	(118.63)	(115.45)
	Panel C: Log Annual Earnings			
Nonattainment×1($\tau > 1971$)	0.006	0.010**	0.010**	0.011**
	(0.005)	(0.005)	(0.004)	(0.004)
	Panel D: Non-Zero Earnings			
Nonattainment×1($\tau > 1971$)	45.05	95.89	95.55	111.45
	(99.66)	(93.06)	(83.68)	(84.82)
Counties	148	148	148	148
First Year	1969	1969	1969	1969
Last Year	1974	1974	1974	1974
Sample Size	888	888	888	888
Year of Birth Weather Controls		Yes	Yes	Yes
Year of Birth Natality Basic			Yes	Yes
Year of Birth Natality Unrestricted				Yes

NOTE: This table reports regression coefficients from 16 separate regressions, 4 per panel. The dependent variable in each regression is a composition-adjusted labor market average for a birth-county×birth-year cell. The outcome in Panel A is the mean annual quarters employed $\in [0, 4]$ between ages 29 and 31, the outcome in Panel B is mean annual earnings between ages 29 and 31, the outcome in Panel C is the mean annual log earnings between ages 29 and 31, and the outcome in Panel D is the mean annual non-zero earnings between ages 29 and 31. The key explanatory variable in the regression model is Nonattainment×1($\tau > 1971$), which is an indicator equal to one for counties designated as nonattainment interacted with an indicator equal to one for birth years after 1971. The regressions control for birth-county fixed effects, birth-state×birth-year fixed effects, and predetermined county characteristics (population, employment, total transfers) interacted with quadratic trends. Additional birth-county time-varying controls are listed in the table and described further in the text. The regressions are weighted by the birth-county×birth-year cell size. Standard errors are clustered by commuting zone and are in parentheses. See text for details. Source: LEHD, EPA.

Table 4: The Effect of TSP Exposure in the Year of Birth on Long-Run Labor Market Outcomes at Ages 29-31, 2SLS Estimates

	(1)	(2)	(3)	(4)
	Panel A: Quarters Employed			
Mean TSPs (/10)	-0.017*	-0.024**	-0.021*	-0.028*
	(0.010)	(0.012)	(0.011)	(0.016)
	Panel B: Annual Earnings			
Mean TSPs (/10)	-186.21	-286.83*	-266.88*	-351.74*
	(139.30)	(158.61)	(145.62)	(190.39)
	Panel C: Log Annual Earnings			
Mean TSPs (/10)	-0.005	-0.010*	-0.011*	-0.014*
	(0.005)	(0.006)	(0.006)	(0.008)
	Panel D: Non-Zero Earnings			
Mean TSPs (/10)	-37.35	-98.67	-101.76	-141.69
	(83.22)	(97.97)	(89.03)	(107.43)
First Stage F-Statistic	14.7	15.0	14.8	14.2
Counties	148	148	148	148
First Year	1969	1969	1969	1969
Last Year	1974	1974	1974	1974
Sample Size	888	888	888	888
Year of Birth Weather Controls		Yes	Yes	Yes
Year of Birth Natality Basic			Yes	Yes
Year of Birth Natality Unrestricted				Yes

NOTE: This table reports regression coefficients from 16 separate regressions, 4 per panel. The dependent variable in each regression is a composition-adjusted labor market average for a birth-county×birth-year cell. The outcome in Panel A is the mean annual quarters employed $\in [0, 4]$ between ages 29 and 31, the outcome in Panel B is mean annual earnings between ages 29 and 31, the outcome in Panel C is the mean annual log earnings between ages 29 and 31, and the outcome in Panel D is the mean annual non-zero earnings between ages 29 and 31. The regression models estimate the effects of the TSP level in the year of birth, using Nonattainment$\times 1(\tau > 1971)$ as an instrumental variable. Kleibergen-Paap Wald F statistics are reported as a test of the first-stage strength of the instrument. The regression models control for birth-county fixed effects, birth-state×birth-year fixed effects, and predetermined county characteristics (population, employment, total transfers) interacted with quadratic trends. Additional birth-county time-varying controls are listed in the table and described further in the text. The regressions are weighted by the birth-county×birth-year cell size. Standard errors are clustered by commuting zone and are in parentheses. See text for details. Source: LEHD, EPA.

Table 5: The Effect of CAAA Implementation on Long-Run Labor Market Outcomes by Age of Observation

	(1) Age 27	(2) Age 28	(3) Age 29	(4) Age 30	(5) Age 31	(6) Age 32	(7) Age 33	(8) Age 34	(9) Age 35	(10) Age 27-35
	Panel A: Quarters Employed									
Nonattainment×1($\tau > 1971$)	0.028***	0.023***	0.022***	0.022**	0.021***	0.015**	0.015*	0.014	0.012	0.018**
	(0.009)	(0.008)	(0.008)	(0.009)	(0.008)	(0.006)	(0.009)	(0.009)	(0.009)	(0.009)
	Panel B: Annual Earnings									
Nonattainment×1($\tau > 1971$)	211.27*	97.86	209.35**	286.41**	332.94**	219.31**	198.13	185.57	200.39	255.92**
	(113.77)	(111.47)	(105.70)	(132.00)	(130.48)	(104.09)	(104.57)	(107.79)	(124.31)	(116.39)
Counties	148	148	148	148	148	148	148	148	148	148
First Year	1971	1970	1969	1969	1969	1969	1969	1969	1969	1969
Last Year	1974	1974	1974	1974	1974	1974	1974	1973	1972	1974
Sample Size	592	740	888	888	888	888	888	740	592	888
YOB Weather Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
YOB Natality Basic	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
YOB Natality Unrestricted	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

NOTE: This table reports regression coefficients from 20 separate regressions, 10 per panel. The dependent variable in each regression is a composition-adjusted labor market average for a birth-county×birth-year cell. The outcome in Panel A is the mean annual quarters employed, and the outcome in Panel B is mean annual earnings. The outcome variables in both panels change with respect to follow-up ages (indicated in the column headings). The regression models estimate the effects nonattainment designation for cohorts born in the years after the CAAA went into effect. All regressions control for birth-county fixed effects, birth-state×birth-year fixed effects, and predetermined county characteristics (population, employment, total transfers) interacted with linear trends. Additional birth-county time-varying controls are listed in the table and described further in the text. The regressions are weighted by the birth-county×birth-year cell size. Standard errors are clustered by commuting zone and are in parentheses. See text for details. Source: LEHD, EPA.

Table 6: The Effect of TSP Exposure in the Year of Birth on Labor Market Outcomes by Age of Exposure, 2SLS Estimates

	(1) Age 27	(2) Age 28	(3) Age 29	(4) Age 30	(5) Age 31	(6) Age 32	(7) Age 33	(8) Age 34	(9) Age 35	(10) Age 27-35
	Panel A: Quarters Employed									
Mean TSPs (/10)	-0.026**	-0.025*	-0.028*	-0.028*	-0.026*	-0.019	-0.019	-0.015	-0.021*	-0.020**
	(0.011)	(0.013)	(0.014)	(0.016)	(0.015)	(0.012)	(0.012)	(0.012)	(0.012)	(0.011)
	Panel B: Annual Earnings									
Mean TSPs (/10)	-198.81*	-105.52	-260.61*	-356.50*	-414.17*	-272.77*	-246.51	-214.31	-202.89	-280.22**
	(120.11)	(121.43)	(149.69)	(204.88)	(221.57)	(160.41)	(160.40)	(178.84)	(190.67)	(142.73)
First Stage F-Statistic	30.5	15	14.2	14.2	14.2	14.2	14.2	5.7	2.3	15.1
Counties	148	148	148	148	148	148	148	148	148	148
First Year	1971	1970	1969	1969	1969	1969	1969	1969	1969	1969
Last Year	1974	1974	1974	1974	1974	1974	1974	1973	1972	1974
Sample Size	592	740	888	888	888	888	888	740	592	888
YOB Weather Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
YOB Natality Basic	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
YOB Natality Unrestricted	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

NOTE: This table reports regression coefficients from 20 separate regressions, 10 per panel. The dependent variable in each regression is a composition-adjusted labor market average for a birth-county×birth-year cell. The outcome in Panel A is the mean annual quarters employed, and the outcome in Panel B is mean annual earnings. The outcome variables in both panels change with respect to follow-up ages (indicated in the column headings). The table reports results from estimating the effects of the TSP level in the year of birth, using Nonattainment$\times 1(\tau > 1971)$ as an instrumental variable. Kleibergen-Paap Wald F statistics are reported as a test of the first-stage strength of the instrument. All regressions control for birth-county fixed effects, birth-state×birth-year fixed effects, and predetermined county characteristics (population, employment, total transfers) interacted with linear trends. Additional birth-county time-varying controls are listed in the table. The regressions are weighted by the birth-county×birth-year cell size. Standard errors are clustered by commuting zone and are in parentheses. See text for details. Source: LEHD, EPA.

Table 7: Heterogeneous Effects of CAAA Implementation Across the Earnings Distribution

	(1) $0 \leq p \leq 1$	(2) $1 < p \leq 5$	(3) $5 < p \leq 10$	(4) $10 < p \leq 25$	(5) $25 < p \leq 50$
Nonattainment$\times 1(\tau > 1971)$	-0.007***	-0.000	-0.000	0.002***	0.003**
	(0.002)	(0.000)	(0.000)	(0.001)	(0.001)
Average Income Cutoffs	$[0, 4324]$	$(4324, 9873]$	$(9873, 14116]$	$(14116, 22983]$	$(22983, 34721]$
	(6) $50 \leq p \leq 75$	(7) $75 < p \leq 90$	(8) $90 < p \leq 95$	(9) $95 < p \leq 99$	(10) $99 < p \leq 100$
Nonattainment$\times 1(\tau > 1971)$	0.001	0.002**	-0.000	0.000	-0.000
	(0.001)	(0.001)	(0.000)	(0.000)	(0.000)
Average Income Cutoffs	$(34721, 48357]$	$(48357, 63949]$	$(63949, 74942]$	$(74942, 93957]$	$(93957, 100000]$
Counties	148	148	148	148	148
First Year	1969	1969	1969	1969	1969
Last Year	1974	1974	1974	1974	1974
Sample Size	888	888	888	888	888
YOB Weather Controls	Yes	Yes	Yes	Yes	Yes
YOB Natality Basic	Yes	Yes	Yes	Yes	Yes
YOB Natality Unrestricted	Yes	Yes	Yes	Yes	Yes

NOTE: This table reports regression coefficients from 10 separate regressions, 5 per panel. The dependent variable in each column is the proportion of individuals in a birth-county×birth-year cohort whose age 29-31 earnings falls within the 1969 ("pre-treatment") earnings quantile indicated in the column heading. The table also reports the average income cutoffs pertaining to the various percentiles of the pre-treatment earnings distribution. The table reports estimates from reduced form models, estimating the effects nonattainment designation for cohorts born in the years after the CAAA went into effect. All regressions control for birth-county fixed effects, state×year fixed effects, and predetermined county characteristics (population, employment, total transfers) interacted with quadratic trends. Additional birth-county time-varying controls are listed in the table. Regressions are weighted by the birth-county×birth-year cell size. Standard errors are clustered by commuting zone and are in parentheses. Source: LEHD, EPA.

Table 8: The Effect of CAAA Implementation on Earnings at Ages 29-31, by Race and Gender

	(1) White	(2) African American	(3) Asian	(4) Hispanic	(5) Male	(6) Female
Nonattainment×1($\tau > 1971$)	382.49**	208.18**	209.40*	225.72**	288.90**	156.19
	(150.46)	(95.88)	(107.09)	(107.36)	(116.36)	(126.05)
Nonattainment×1($\tau > 1971$)×Group	-162.43	11.15	133.63	602.36*	-128.35	117.09
	(133.81)	(133.33)	(135.01)	(312.29)	(125.32)	(123.19)
Counties	148	148	148	148	148	148
First Year	1969	1969	1969	1969	1969	1969
Last Year	1974	1974	1974	1974	1974	1974
Sample Size	5380119	5380119	5380119	5380119	5380119	5380119
YOB Weather Controls	Yes	Yes	Yes	Yes	Yes	Yes
YOB Natality Basic	Yes	Yes	Yes	Yes	Yes	Yes
YOB Natality Unrestricted	Yes	Yes	Yes	Yes	Yes	Yes

NOTE: This table reports regression coefficients from 6 separate regressions. The regressions are estimated using the individual-level micro data. The dependent variable in all columns is mean annual annual earnings between ages 29 and 31, and the regression sample remains the same across all columns. Both panels report estimates from reduced form models that include additional interaction terms between our main treatment variable, Nonattainment×1($\tau > 1971$), and a group-specific dummy variable indicated in the column heading. We also include the additional lower-order interaction terms (i.e., Race/Sex×Nonattainment and Race/Sex×1($\tau > 1971$)) in all regression models. All regressions control for birth-county fixed effects, birth-state×birth-year fixed effects, and predetermined county characteristics (population, employment, total transfers) interacted with quadratic trends. Additional birth-county time-varying controls are listed in the table and described further in the text. The regressions are unweighted. Standard errors are clustered by commuting zone and are in parentheses. See text for details. Source: LEHD, EPA.

Table 9: The Effect of CAAA Implementation on Quarters Worked at Ages 29-31, by Race and Gender

	(1) White	(2) African American	(3) Asian	(4) Hispanic	(5) Male	(6) Female
Nonattainment×1($\tau > 1971$)	0.013	0.020***	0.017**	0.019***	0.018**	0.019***
	(0.011)	(0.006)	(0.006)	(0.007)	(0.008)	(0.007)
Nonattainment×1($\tau > 1971$)×Group	0.009	-0.017*	0.003	-0.029	0.001	-0.002
	(0.013)	(0.010)	(0.010)	(0.025)	(0.007)	(0.007)
Counties	148	148	148	148	148	148
First Year	1969	1969	1969	1969	1969	1969
Last Year	1974	1974	1974	1974	1974	1974
Sample Size	5380119	5380119	5380119	5380119	5380119	5380119
YOB Weather Controls	Yes	Yes	Yes	Yes	Yes	Yes
YOB Natality Basic	Yes	Yes	Yes	Yes	Yes	Yes
YOB Natality Unrestricted	Yes	Yes	Yes	Yes	Yes	Yes

NOTE: This table reports regression coefficients from 6 separate regressions. The regressions are estimated using the individual-level micro data. The dependent variable in all columns is mean annual quarters worked from ages 29-31, and the regression sample remains the same across all columns. Both panels report estimates from reduced form models that include additional interaction terms between our main treatment variable, Nonattainment×1($\tau > 1971$), and a group-specific dummy variable indicated in the column heading. We also include the additional lower-order interaction terms (i.e., Race/Sex×Nonattainment and Race/Sex×1($\tau > 1971$)) in all regression models. All regressions control for birth-county fixed effects, birth-state×birth-year fixed effects, and predetermined county characteristics (population, employment, total transfers) interacted with quadratic trends. Additional birth-county time-varying controls are listed in the table and described further in the text. The regressions are unweighted. Standard errors are clustered by commuting zone and are in parentheses. See text for details. Source: LEHD, EPA.

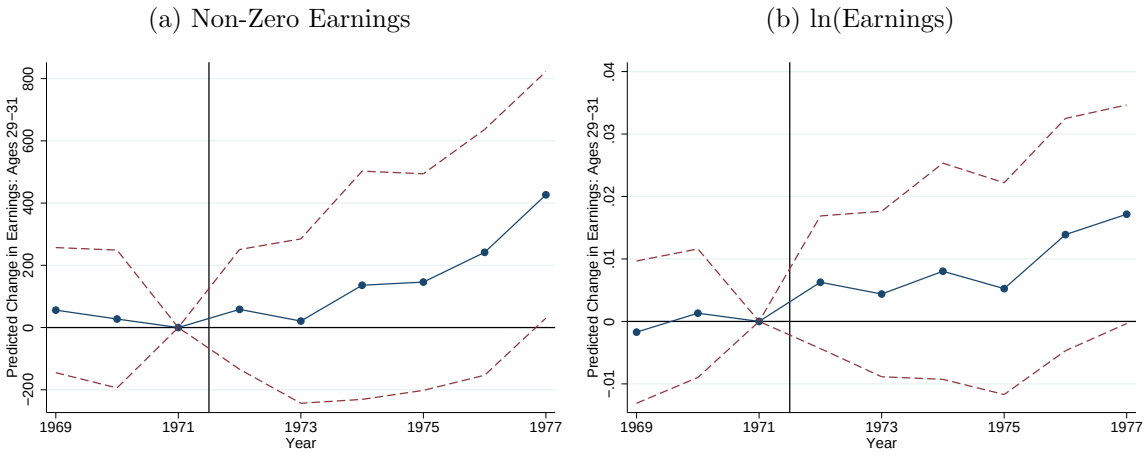
Table 10: The Effect of CAAA Implementation on Sorting and Population Characteristics

	(1) Maternal Education	(2) Fraction White	(3) Fraction Black	(4) Income Per Capita	(5) Predicted Earnings Ages 29-31
Nonattainment×1($\tau > 1971$)	-0.155	0.007	-0.004	-37.23	-24.58
	(0.099)	(0.005)	(0.003)	(106.20)	(29.82)
Variable Mean	12	0.839	0.140	18506	23587
Counties	101	148	148	148	148
First Year	1969	1969	1969	1969	1969
Last Year	1974	1974	1974	1974	1974
Sample Size	606	888	888	888	888
Year of Birth Weather Controls	Yes	Yes	Yes	Yes	Yes
Year of Birth Natality Basic					
Year of Birth Natality Unrestricted					

NOTE: This table reports regression coefficients from 5 separate regressions. The dependent variables in columns (1)-(4) are population-level average characteristics in each birth-county×birth-year cell. Column (5) is a predicted earnings measure, where we use our individual-level micro data from the LEHD and regress earnings at ages 29-31 on a set of indicators for race and sex. The predicted values from this regression are used as a summary index measure of population characteristics (averaged to the birth-county×year level). The regression sample remains the same in all columns. The table reports estimates from reduced form models, estimating the effects nonattainment designation for cohorts born in the years after the CAAA went into effect. All regressions control for birth-county fixed effects and birth-state×birth-year fixed effects. The regressions are weighted by birth-county×birth-year cell size. Standard errors are clustered by commuting zone and are in parentheses. See text for details. Source: LEHD, NCHS, EPA.

A Appendix Figures and Tables

Figure A1: The Effect of CAAA Implementation on Long-Run Labor Market Outcomes at Ages 29-31 by Year of Birth, Intensive Margin



NOTE: This figure presents regression coefficients from a version of equation (6). The dependent variables are indicated in the figure sub-headings. We regress the dependent variable on a full set of birth-year indicators interacted with an indicator for birth-county nonattainment status, birth-county fixed effects, birth-state×birth-year fixed effects, birth-county economic controls (population, employment, total transfers), weather controls, and unrestricted natality controls, weighting by the birth-county×birth-year cell size. We plot the coefficients on the interactions between birth-year indicators and nonattainment. The coefficient on the interaction between birth-year 1971 and nonattainment is normalized to zero. The dashed lines represent 95% confidence intervals, where standard errors are clustered by commuting zone. Source: LEHD, EPA.

Table A1: Summary Means in 1969 and Pre-Trend Differences Between 1969-1971, Above/Below Median Pollution Levels

	(1) Below Median TSP	(2) Above Median TSP	(3) (2)-(1) p-value
Cohort Averages: Ages 29-31 (Source: LEHD)			
Earnings (2008$)	23648	24293	0.943
Earnings: 4-Quarter (2008$)	37298	38021	0.517
Quarters Employed	2.741	2.75	0.061
Fraction Working in Non-Birth State	0.295	0.309	0.261
County Environment: Age 0 (Source: EPA, Schlenker and Roberts (2009))			
Total Suspended Particulate ($\mu g/m^3$)	54.5	96.32	0.000
Average Annual Precipitation	1222	906	0.000
Average Temperature	12.86	13.19	0.993
County Socio-Economic: Age 0 (Source: BEA)			
Income Per Capita (2008$)	17237	18010	0.528
Total Employment	105372	253430	0.006
Transfers UI Per Capita	4.466	4.366	0.530
Total Transfers Per Capita	7.396	7.351	0.592
Cohort Demographics: Age 0 (Source: NCHS)			
% White	0.858	0.841	0.022
% African American	0.109	0.147	0.003
Mother's Education	11.91	11.82	0.131
Father's Education	12.31	12.38	0.359
Mother's Age	24.11	24.39	0.187
Father's Age	27.34	27.49	0.203

NOTE: This is a table of summary means from our baseline estimation sample. The summary statistics are calculated using data from 1969 for a single cross-section of counties. An observation is a county, and the means have been weighted by county population. All dollar amounts are in 2008 dollars. Column (1) presents summary statistics for those counties that were below the median level of county-level TSP readings in 1970. Column (2) presents summary statistics for the subset of counties that were above the median level of county-level TSP readings in 1970. Column (3) presents the p-values from a test of the null hypothesis that the levels in columns (1) and (2) are the same in 1969. The p-values in Columns (3) are generated from a regression of levels on an indicator for above median TSP, using robust standard errors for inference, and weighting by the county population.

Table A2: Summary Means in 1969 and Pre-Trend Differences Between 1969-1971, Above/Below Median Pollution Changes

	(1) Below Median ΔTSP	(2) Above Median ΔTSP	(3) (2)-(1) p-value
Cohort Averages: Ages 29-31 (Source: LEHD)			
Earnings (2008$)	23757	24464	0.323
Earnings: 4-Quarter (2008$)	37613	37865	0.521
Quarters Employed	2.729	2.777	0.485
Fraction Working in Non-Birth State	0.302	0.303	0.152
County Environment: Age 0 (Source: EPA, Schlenker and Roberts (2009))			
Total Suspended Particulate ($\mu g/m^3$)	71.52	89.4	0.037
Average Annual Precipitation	1037	1061	0.756
Average Temperature	12.73	13.6	0.309
County Socio-Economic: Age 0 (Source: BEA)			
Income Per Capita (2008$)	17525	17930	0.881
Total Employment	170759	220753	0.793
Transfers UI Per Capita	4.379	4.465	0.481
Total Transfers Per Capita	7.384	7.346	0.197
Cohort Demographics: Age 0 (Source: NCHS)			
% White	0.848	0.850	0.546
% African American	0.127	0.136	0.823
Mother's Education	11.98	11.64	0.000
Father's Education	12.45	12.17	0.000
Mother's Age	24.15	24.48	0.016
Father's Age	27.3	27.66	0.070

NOTE: This is a table of summary means from our baseline estimation sample. The summary statistics are calculated using data from 1969 for a single cross-section of counties. An observation is a county, and the means have been weighted by county population. All dollar amounts are in 2008 dollars. Column (1) presents summary statistics for those counties that experience below median TSP changes between 1970 and 1972. Column (2) presents summary statistics for the subset of counties that experienced above median TSP changes between 1970 and 1972. Column (3) presents the p-values from a test of the null hypothesis that the levels in columns (1) and (2) are the same in 1969. The p-values in Column (3) are generated from a regression of levels on an indicator for above median TSP changes, using robust standard errors for inference, and weighting by the county population.

Table A3: The Effect of CAAA Implementation on Ambient TSP Exposure

	(1)	(2)	(3)	(4)
Nonattainment×1($\tau > 1971$)	-12.06***	-9.71***	-9.38***	-7.86***
	(3.31)	(2.78)	(2.56)	(2.84)
Counties	148	148	148	148
First Year	1969	1969	1969	1969
Last Year	1974	1974	1974	1974
Sample Size	888	888	888	888
Weather Controls		Yes	Yes	Yes
Natality Basic			Yes	Yes
Natality Unrestricted				Yes

NOTE: This table reports regression coefficients from 4 separate regressions. The dependent variable the annual mean TSP reading in a county×year. The regression sample remains the same in all columns. The table reports OLS estimates from a first stage regression, estimating the effect of nonattainment designation on ambient TSP levels in the years after the CAAA. Nonattainment×1($\tau > 1971$) is an indicator variable equal to one for counties designated as nonattainment based on 1970 ambient TSP levels in the years after the regulations go into place. All regressions control for county fixed effects, state×year fixed effects, and predetermined county characteristics (population, employment, total tranfers) interacted with quadratic trends. Additional controls are listed in the tables. Regressions are weighted by the birth-county×birth-year cell size. Standard errors are clustered by commuting zone and are in parentheses. See text for details. Source: EPA.

Table A4: The Effects of CAAA Implementation and TSP Exposure in the Year of Birth on Earnings at Ages 29-31, Individual Micro-Data

	(1)	(2)	(3)	(4)
	Panel A: Nonattainment and Annual Earnings			
Nonattainment×1($\tau > 1971$)	235.06	286.40**	263.86**	285.09**
	(146.00)	(125.30)	(115.19)	(111.42)
	Panel B: TSPs and Annual Earnings			
Mean TSPs (/10)	-194.94	-294.75*	-281.01*	-362.58*
	(139.32)	(161.91)	(146.06)	(191.63)
First Stage F-Statistic	17.4	18.2	18.2	18.2
Counties	148	148	148	148
First Year	1969	1969	1969	1969
Last Year	1974	1974	1974	1974
Sample Size	5380119	5380119	5380119	5380119
Year of Birth Weather Controls		Yes	Yes	Yes
Year of Birth Natality Basic			Yes	Yes
Year of Birth Natality Unrestricted				Yes

NOTE: This table reports regression coefficients from 8 separate regressions, 4 per panel. The regressions are estimated using the individual-level micro data. The dependent variable in all columns is mean annual earnings between ages 29 and 31, and the regression sample remains the same across all columns. Panel A reports the reduced form results from estimating the effects nonattainment designation for cohorts born in the years after the CAAA went into effect. The key explanatory variable in Panel A is Nonattainment×1($\tau > 1971$), which is an indicator equal to one for counties designated as nonattainment interacted with an indicator equal to one for birth years after 1971. Panel B reports the 2SLS results from estimating the effects of the TSP level in the year of birth, using Nonattainment×1($\tau > 1971$) as an instrumental variable. Kleibergen-Paap Wald F statistics are reported as a test of the first-stage strength of the instrument. All regressions control for individual-level covariates (indicators for race, sex, and month of birth), a vector of contemporaneous state-level time-varying controls observed in the individual's state of work. All regressions also include birth-county fixed effects, birth-state×birth-year fixed effects, and predetermined county characteristics (population, employment, total transfers) interacted with quadratic trends. Additional birth-county time-varying controls are listed in the table and described further in the text. The regressions are unweighted. Standard errors are clustered by commuting zone and are in parentheses. See text for details. Source: LEHD, EPA.

Table A5: The Effects of CAAA Implementation and TSP Exposure in the Year of Birth on Quarters Worked at Ages 29-31, Individual Micro-Data

	(1)	(2)	(3)	(4)
	Panel A: Nonattainment and Quarters Employed			
Nonattainment$\times 1(\tau > 1971)$	0.019**	0.021***	0.018**	0.021***
	(0.009)	(0.008)	(0.008)	(0.007)
	Panel B: TSPs and Quarters Employed			
Mean TSPs (/10)	-0.016	-0.022*	-0.020*	-0.027*
	(0.010)	(0.011)	(0.010)	(0.015)
First Stage F-Statistic	17.4	18.2	18.2	18.2
Counties	148	148	148	148
First Year	1969	1969	1969	1969
Last Year	1974	1974	1974	1974
Sample Size	5380119	5380119	5380119	5380119
Year of Birth Weather Controls		Yes	Yes	Yes
Year of Birth Natality Basic			Yes	Yes
Year of Birth Natality Unrestricted				Yes

NOTE: This table reports regression coefficients from 8 separate regressions, 4 per panel. The regressions are estimated using the individual-level micro data. The dependent variable in all columns is the mean annual quarters employed from ages 29-31, and the regression sample remains the same across all columns. Panel A reports the reduced form results from estimating the effects nonattainment designation for cohorts born in the years after the CAAA went into effect. The key explanatory variable in Panel A is Nonattainment$\times 1(\tau > 1971)$, which is an indicator equal to one for counties designated as nonattainment interacted with an indicator equal to one for birth years after 1971. Panel B reports the 2SLS results from estimating the effects of the TSP level in the year of birth, using Nonattainment$\times 1(\tau > 1971)$ as an instrumental variable. Kleibergen-Paap Wald F statistics are reported as a test of the first-stage strength of the instrument. All regressions control for individual-level covariates (indicators for race, sex, and month of birth), a vector of contemporaneous state-level time-varying controls observed in the individual's state of work. All regressions also include birth-county fixed effects, birth-state$\times$birth-year fixed effects, and predetermined county characteristics (population, employment, total transfers) interacted with quadratic trends. Additional birth-county time-varying controls are listed in the table and described further in the text. The regressions are unweighted. Standard errors are clustered by commuting zone and are in parentheses. See text for details. Source: LEHD, EPA.

Table A6: The Effect of CAAA Implementation on Labor Market Outcomes at Ages 29-31, Shorter Time Window

	(1)	(2)	(3)	(4)
	Panel A: Quarters Employed			
Nonattainment×1($\tau > 1971$)	0.020***	0.019**	0.015*	0.018**
	(0.007)	(0.008)	(0.008)	(0.008)
	Panel B: Annual Earnings			
Nonattainment×1($\tau > 1971$)	168.77*	210.91**	187.85*	171.90*
	(97.85)	(95.42)	(95.85)	(102.41)
	Panel C: Log Annual Earnings			
Nonattainment×1($\tau > 1971$)	0.004	0.007	0.005	0.007
	(0.005)	(0.005)	(0.005)	(0.005)
	Panel D: Non-Zero Earnings			
Nonattainment×1($\tau > 1971$)	2.46	68.80	61.50	31.62
	(89.24)	(91.38)	(82.13)	(91.91)
Counties	148	148	148	148
First Year	1970	1970	1970	1970
Last Year	1972	1972	1972	1972
Sample Size	444	444	444	444
Year of Birth Weather Controls		Yes	Yes	Yes
Year of Birth Natality Basic			Yes	Yes
Year of Birth Natality Unrestricted				Yes

NOTE: This table reports regression coefficients from 16 separate regressions, 4 per panel. The dependent variable in each regression is a composition-adjusted labor market average for a birth-county×birth-year cell. The outcome in Panel A is the mean annual quarters employed $\in [0, 4]$ between ages 29 and 31, the outcome in Panel B is mean annual earnings between ages 29 and 31, the outcome in Panel C is the mean annual log earnings between ages 29 and 31, and the outcome in Panel D is the mean annual non-zero earnings between ages 29 and 31. The key explanatory variable in the regression model is Nonattainment×1($\tau > 1971$), which is an indicator equal to one for counties designated as nonattainment interacted with an indicator equal to one for birth years after 1971. The regressions control for birth-county fixed effects, birth-state×birth-year fixed effects, and predetermined county characteristics (population, employment, total transfers) interacted with quadratic trends. Additional birth-county time-varying controls are listed in the table and described further in the text. The regressions are weighted by the birth-county×birth-year cell size. Standard errors are clustered by commuting zone and are in parentheses. See text for details. Source: LEHD, EPA.

Table A7: The Effect of TSP Exposure in the Year of Birth on Labor Market Outcomes at Ages 29-31, 2SLS Estimates – Shorter Time Window

	(1)	(2)	(3)	(4)
	Panel A: Quarters Employed			
Mean TSPs (/10)	-0.019**	-0.029	-0.027	-0.033
	(0.009)	(0.022)	(0.024)	(0.035)
	Panel B: Annual Earnings			
Mean TSPs (/10)	-165.99*	-327.65	-332.94	-326.46
	(100.50)	(230.84)	(265.17)	(334.88)
	Panel C: Log Annual Earnings			
Mean TSPs (/10)	-0.003	-0.010	-0.009	-0.012
	(0.005)	(0.009)	(0.010)	(0.014)
	Panel D: Non-Zero Earnings			
Mean TSPs (/10)	-2.42	-106.88	-109.00	-60.05
	(87.64)	(139.42)	(146.92)	(172.63)
First Stage F-Statistic	7.94	9.78	9.52	6.47
Counties	148	148	148	148
First Year	1970	1970	1970	1970
Last Year	1972	1972	1972	1972
Sample Size	444	444	444	444
Year of Birth Weather Controls		Yes	Yes	Yes
Year of Birth Natality Basic			Yes	Yes
Year of Birth Natality Unrestricted				Yes

NOTE: This table reports regression coefficients from 16 separate regressions, 4 per panel. The dependent variable in each regression is a composition-adjusted labor market average for a birth-county×birth-year cell. The outcome in Panel A is the mean annual quarters employed $\in [0, 4]$ between ages 29 and 31, the outcome in Panel B is mean annual earnings between ages 29 and 31, the outcome in Panel C is the mean annual log earnings between ages 29 and 31, and the outcome in Panel D is the mean annual non-zero earnings between ages 29 and 31. The regression models estimate the effects of the TSP level in the year of birth, using Nonattainment$\times 1(\tau > 1971)$ as an instrumental variable. Kleibergen-Paap Wald F statistics are reported as a test of the first-stage strength of the instrument. The regression models control for birth-county fixed effects, birth-state×birth-year fixed effects, and predetermined county characteristics (population, employment, total transfers) interacted with quadratic trends. Additional birth-county time-varying controls are listed in the table and described further in the text. The regressions are weighted by the birth-county×birth-year cell size. Standard errors are clustered by commuting zone and are in parentheses. See text for details. Source: LEHD, EPA.

Table A8: The Effects of CAAA Implementation and TSP Exposure in the Year of Birth on Fertility and Workforce Composition

	(1) Log Number of Workers	(2) Log Number of Births	(3) Ratio M/F Births	(4) Ratio M/F Workers	(5) Ratio Births/Workers
	Panel A: Nonattainment and the Workforce/Fertility				
Nonattainment$\times 1(\tau > 1971)$	0.029*	-0.006	0.002	0.001	0.086*
	(0.015)	(0.016)	(0.002)	(0.001)	(0.045)
	Panel B: TSPs and the Workforce/Fertility				
Mean TSPs (/10)	-0.017	0.003	-0.002	-0.001	-0.050
	(0.012)	(0.009)	(0.003)	(0.002)	(0.034)
Variable Mean	7.96	7.89	.513	.506	1.31
First Stage F-Statistic	4.64	4.64	14.2	14.2	4.64
Counties	148	148	148	148	148
First Year	1969	1969	1969	1969	1969
Last Year	1974	1974	1974	1974	1974
Sample Size	888	888	888	888	888
Weather Controls	Yes	Yes	Yes	Yes	Yes
Natality Basic	Yes	Yes	Yes	Yes	Yes
Natality Unrestricted	Yes	Yes	Yes	Yes	Yes

NOTE: This table reports regression coefficients from 10 separate regressions, 5 per panel. The dependent variable in each column is corresponds to a different outcome in a birth-county×birth-year. The regression sample remains the same in all columns and panels. Panel A reports estimates from reduced form models, estimating the effects nonattainment designation for cohorts born in the years after the CAAA went into effect. Panel B reports the 2SLS results from estimating the effects of the TSP level in the year of birth, using Nonattainment$\times 1(\tau > 1971)$ as an instrumental variable. Kleibergen-Paap Wald F statistics are reported as a test of the first-stage strength of the instrument. All regressions control for birth-county fixed effects, birth-state×birth-year fixed effects, and predetermined county characteristics (population, employment, total transfers) interacted with linear trends. Additional birth-county time-varying controls are listed in the table. The regressions in columns (3)-(4) are weighted by the birth-county×birth-year cell size. Standard errors are clustered by commuting zone and are in parentheses. See text for details. Source: NCHS, EPA.

Table A9: The Effect of CAAA Implementation on Moving to Out Of Sample States

	(1)	(2)	(3)	(4)
	Panel A: Fraction Working in LEHD State Outside Baseline Sample States			
Nonattainment×1($\tau > 1971$)	0.000	-0.000	0.000	0.000
	(0.001)	(0.001)	(0.001)	(0.001)
	Panel B: Fraction Working in LEHD State Different From Birth State			
Nonattainment×1($\tau > 1971$)	0.002	0.001	0.001	0.002
	(0.002)	(0.002)	(0.002)	(0.002)
Counties	148	148	148	148
First Year	1969	1969	1969	1969
Last Year	1974	1974	1974	1974
Sample Size	888	888	888	888
Year of Birth Weather Controls		Yes	Yes	Yes
Year of Birth Natality Basic			Yes	Yes
Year of Birth Natality Unrestricted				Yes

NOTE: This table reports regression coefficients from 4 separate regressions. The dependent variable in Panel A consists of the fraction of individuals who have worked by 2007 in one of the six states in the LEHD that are not in the main estimation sample of states. The dependent variable in Panel B consists of the fraction of individuals who are working at ages 29-31 in a state other than the state in which they were born. The regression sample remains the same in all columns and panels. The table reports estimates from reduced form models, estimating the effects nonattainment designation for cohorts born in the years after the CAAA went into effect. All regressions control for birth-county fixed effects, birth-state×birth-year fixed effects, and predetermined county characteristics (population, employment, total transfers) interacted with quadratic trends. Additional birth-county time-varying controls are listed in the table. The regressions are weighted by the birth-county×birth-year cell size. Standard errors are clustered by commuting zone and are in parentheses. See text for details. Source: LEHD, EPA.

Table A10: Regression Discontinuity Design, Continuity in Observables Across Regulatory Threshold

	(1) Nonattain	(2) Attainment	(3) Nonattain (1969-1971)	(4) Attainment (1969-1971)	(5) (2)-(1) p-value	(6) (4)-(3) p-value
Cohort Averages: Ages 29-31 (Source: LEHD)						
Earnings (2008$)	23623	23294	-0.024	0.027	0.867	0.797
Earnings: 4-Quarter (2008$)	37109	36654	0.022	0.037	0.697	0.889
Quarters Employed	2.74	2.741	-0.038	-0.017	0.095	0.744
Fraction Working in Non-Birth State	0.31	0.30	-0.007	-0.030	0.145	0.707
County Environment: Age 0 (Source: EPA, Schlenker and Roberts (2009))						
Total Suspended Particulate ($\mu g/m^3$)	95.89	58.59	-0.058	0.044	0.000	0.040
Average Annual Precipitation	988.1	1164	-0.034	-0.041	0.991	0.623
Average Temperature	12.51	12.78	0.020	0.007	0.936	0.701
County Socio-Economic: Age 0 (Source: BEA)						
Income Per Capita (2008$)	17181	16613	0.037	0.035	0.976	0.885
Total Employment	194709	67169	0.025	0.034	0.835	0.537
Transfers UI Per Capita	4.374	4.33	0.243	0.224	0.642	0.827
Total Transfers Per Capita	7.35	7.36	0.031	0.029	0.204	0.467
Cohort Demographics: Age 0 (Source: NCHS)						
% White	0.85	0.86	-0.011	-0.002	0.367	0.496
% African American	0.14	0.11	0.055	-0.002	0.145	0.600
Mother's Education	11.82	11.73	0.001	-0.008	0.861	0.726
Father's Education	12.31	12.2	0.012	-0.016	0.321	0.323
Mother's Age	24.3	24.11	-0.007	-0.009	0.205	0.439
Father's Age	27.45	27.35	-0.011	-0.011	0.593	0.536
Totals						
Workers Born in 1969 working in 1998-2000	749,000	167,000				
Counties	97	51				

NOTE: This is a table of summary means from our baseline estimation sample along with test statistics pertaining to the difference in outcomes across the CAAA regulatory threshold. The summary statistics are calculated using data from 1969 for a single cross-section of counties. An observation is a county, and the means have been weighted by county population. All dollar amounts are in 2008 dollars. Column (1) presents summary statistics for those counties that exceeded the EPA's NAAQS TSP standard based on monitor readings in 1970. Column (2) presents summary statistics for the subset of counties that were in compliance with the NAAQS standards in 1970. Columns (3) and (4) present log differences between 1969 and 1971 for nonattainment and attainment counties, respectively. Column (5) presents the p-values from a test of the null hypothesis that the levels in columns (1) and (2) are the same in 1969, controlling for the 1970 level of TSP in the county. Column (6) presents the p-values from a test of the null hypothesis that the log differences between 1969 and 1971 are the same for both nonattainment and attainment counties, controlling for the 1970 level of TSP in the county. The p-values in Columns (5) and (6) are generated from a regression of either levels or log differences on an indicator for nonattainment and controlling for the 1970 level of TSP in the county, using robust standard errors for inference, and weighting by the county population.

Table A11: The Effect of County Nonattainment Designation on TSP, Earnings, and Quarters Worked: Regression Discontinuity Design

	County TSP			Earnings			Quarters Worked		
	BW-50	BW-100	BW-150	BW-50	BW-100	BW-150	BW-50	BW-100	BW-150
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Nonattain×1($\tau > 1971$)	-6.68**	-7.034***	-8.315***	126.054	243.370	277.542*	0.007	0.013	0.018*
	(3.079)	(2.714)	(2.613)	(177.488)	157.500	149.400	(0.012)	(0.010)	(0.010)
Sample Size	720	858	882	720	858	882	720	858	882

NOTE: This table reports regression coefficients from 9 separate regressions specifications from a regression discontinuity design. The dependent variable in each regression is a composition-adjusted labor market average for a birth-county×birth-year cell. The outcome variable in columns (1)-(3) is the annual mean TSP in a birth-county×birth-year. The outcome variable in columns (4)-(6) consists of mean annual earnings, and the outcome variable in columns (7)-(9) is the mean annual quarters employed. Each column for each dependent variable is estimated on a different subsample corresponding to a different bandwidth, indicated in the column headings. The running variable is the county 1970 pollution level, and it is interacted with an indicator for a nonattainment county, whose 1970 TSP level exceeds the $75\mu g/m^3$ threshold. All regressions estimate the effects nonattainment designation for cohorts born in the years after the CAAA went into effect (the running variable is set to zero for cohorts born before 1972). All regressions control for birth-county fixed effects, birth-state×birth-year fixed effects, and predetermined county characteristics (population, employment, total transfers) interacted with quadratic trends. Additional birth-county time-varying controls for weather and natality characteristics are described further in the text. The regressions are weighted by the birth-county×birth-year cell size. Standard errors are clustered by commuting zone and are in parentheses. See text for details. Source: LEHD, EPA.

Table A12: The Effect of CAAA Implementation on Average State Level Pollution at Ages 29-31

	(1) State TSP Ages 29-31	(2) State PM10 Ages 29-31
Nonattainment$\times 1(\tau > 1971)$	0.053	0.015
	(0.164)	(0.023)
Counties	148	148
First Year	1969	1969
Last Year	1974	1974
Sample Size	888	888
Weather Controls	Yes	Yes
Natality Basic	Yes	Yes
Natality Unrestricted	Yes	Yes

NOTE: This table reports regression coefficients from 2 separate regressions. The outcome in column (1) is the average TSP pollution level in an individual's state of work between ages 29-31. The outcome in column (2) is the average PM-10 pollution level in an individual's state of work between ages 29-31. The regression sample is the same in both columns. The data are aggregated to birth-county×birth-year cells. The table reports estimates from reduced form models, estimating the effects nonattainment designation for cohorts born in the years after the CAAA went into effect. All regressions control for birth-county fixed effects, birth-state×birth-year fixed effects, and predetermined county characteristics (population, employment, total transfers) interacted with quadratic trends. Additional birth-county time-varying controls for weather and natality characteristics are described further in the text. The regressions are weighted by the birth-county×birth-year cell size. Standard errors are clustered by commuting zone and are in parentheses. See text for details. Source: LEHD, EPA.

Table A13: The Effects of CAAA Implementation and TSP Exposure in the Year of Birth on Labor Market Outcomes at Ages 29-31, By Mobility Status

	(1) Earnings Same State	(2) Earnings Different State	(3) Quarters Worked Same State	(4) Quarters Worked Different State
	Panel A: Nonattainment and Labor Market Outcomes			
Nonattainment$\times 1(\tau > 1971)$	291.108**	322.862***	0.017**	0.030***
	(129.672)	(116.397)	(0.008)	(0.008)
	Panel B: TSPs and Labor Market Outcomes			
Mean TSPs (/10)	-385.567*	-375.942	-0.022	-0.034*
	(200.700)	(230.312)	(0.014)	(0.019)
First Stage F-Statistic	17.1	16.6	17.1	16.6
Counties	148	148	148	148
First Year	1969	1969	1969	1969
Last Year	1974	1974	1974	1974
Sample Size	3765631	1614488	3765631	1614488
Weather Controls	Yes	Yes	Yes	Yes
Natality Basic	Yes	Yes	Yes	Yes
Natality Unrestricted	Yes	Yes	Yes	Yes

NOTE: This table reports regression coefficients from 8 separate regressions, 4 per panel. The regressions are estimated using the individual-level micro data. The dependent variable in columns (1) and (2) is mean annual earnings, and the outcome in columns (3) and (4) is the mean annual quarters employed. The regression sample in columns (1) and (3) consists of all individuals who work in their state of birth at age 30, and columns (2) and (4) consists of all individuals who do not work in their state of birth at age 30. Panel A reports estimates from reduced form models, estimating the effects nonattainment designation for cohorts born in the years after the CAAA went into effect. Panel B reports the 2SLS results from estimating the effects of the TSP level in the year of birth, using Nonattainment$\times 1(\tau > 1971)$ as an instrumental variable. Kleibergen-Paap Wald F statistics are reported as a test of the first-stage strength of the instrument. All regressions control for birth-county fixed effects, birth-state$\times$birth-year fixed effects, and predetermined county characteristics (population, employment, total transfers) interacted with quadratic trends. Additional birth-county time-varying controls for weather and natality characteristics are described further in the text. Standard errors are clustered by commuting zone and are in parentheses. See text for details. Source: LEHD, EPA.

Table A14: The Effects of CAAA Implementation and TSP Exposure in the Year of Birth on Infant Mortality and Morbidity

	(1) Birthweight	(2) Low Birthweight	(3) Internal IMR Rate	(4) External IMR Rate
	Panel A: Nonattainment and Infant Health			
Nonattainment$\times 1(\tau > 1971)$	-6.838**	0.0005	-0.0005	0.0002
	(3.009)	(0.0012)	(0.0008)	(0.0002)
	Panel B: TSPs and Infant Health			
Mean TSPs (/10)	8.694*	-0.0006	0.0006	-0.0002
	(4.588)	(0.0016)	(0.0010)	(0.0002)
Variable Mean	3273	0.0784	0.0340	0.0014
First Stage F-Statistic	14.2	14.2	14.2	14.2
Counties	148	148	148	148
First Year	1969	1969	1969	1969
Last Year	1974	1974	1974	1974
Sample Size	888	888	888	888
Weather Controls	Yes	Yes	Yes	Yes
Natality Basic	Yes	Yes	Yes	Yes
Natality Unrestricted	Yes	Yes	Yes	Yes

NOTE: This table reports regression coefficients from 8 separate regressions, 4 per panel. The dependent variable in each column corresponds to a different infant health outcome in a birth-county×birth-year. The regression sample remains the same in all columns and panels. Panel A reports estimates from reduced form models, estimating the effects nonattainment designation for cohorts born in the years after the CAAA went into effect. Panel B reports the 2SLS results from estimating the effects of the TSP level in the year of birth, using Nonattainment$\times 1(\tau > 1971)$ as an instrumental variable. Kleibergen-Paap Wald F statistics are reported as a test of the first-stage strength of the instrument. All regressions control for birth-county fixed effects, birth-state×birth-year fixed effects, and predetermined county characteristics (population, employment, total transfers) interacted with quadratic trends. Additional controls for weather and natality characteristics are listed in the table and described further in the text. The regressions are weighted by the birth-county×birth-year cell size. Standard errors are clustered by commuting zone and are in parentheses. See text for details. Source: NCHS, EPA.

Table A15: The Effects of CAAA Implementation and TSP Exposure in the Year of Birth on Infant Mortality and Morbidity, Shorter Time Window

	(1) Birthweight	(2) Low Birthweight	(3) Internal IMR Rate	(4) External IMR Rate
	Panel A: Nonattainment and Infant Health			
Nonattainment×1($\tau > 1971$)	-4.837	-0.0008	-0.0045***	0.0005*
	(4.549)	(0.0020)	(0.0012)	(0.0003)
	Panel B: TSPs and Infant Health			
Mean TSPs (/10)	9.185	0.0016	0.0086	-0.0010
	(10.951)	(0.0040)	(0.0072)	(0.0010)
Variable Mean	3272	0.0781	0.0358	0.0015
First Stage F-Statistic	6.47	6.47	6.47	6.47
Counties	148	148	148	148
First Year	1970	1970	1970	1970
Last Year	1972	1972	1972	1972
Sample Size	444	444	444	444
Weather Controls	Yes	Yes	Yes	Yes
Natality Basic	Yes	Yes	Yes	Yes
Natality Unrestricted	Yes	Yes	Yes	Yes

NOTE: This table reports regression coefficients from 8 separate regressions, 4 per panel. The sample is limited to cohorts born in 1970, 1971, and 1972. The dependent variable in each column is corresponds to a different infant health outcome in a birth-county×birth-year. The regression sample remains the same in all columns and panels. Panel A reports estimates from reduced form models, estimating the effects nonattainment designation for cohorts born in the years after the CAAA went into effect. Panel B reports the 2SLS results from estimating the effects of the TSP level in the year of birth, using Nonattainment×1($\tau > 1971$) as an instrumental variable. Kleibergen-Paap Wald F statistics are reported as a test of the first-stage strength of the instrument. All regressions control for birth-county fixed effects, birth-state×birth-year fixed effects, and predetermined county characteristics (population, employment, total transfers) interacted with linear trends. Additional controls for weather and natality characteristics are listed in the table and described further in the text. The regressions are weighted by the birth-county×birth-year cell size. Standard errors are clustered by commuting zone and are in parentheses. See text for details. Source: NCHS, EPA.

Table A16: The Effects of CAAA Implementation in the Year of Birth on Labor Market Outcomes at Ages 29-31, Lee (2009) Bounds

	(1) Earnings Lower Bound	(2) Earnings Upper Bound	(3) Quarters Worked Lower Bound	(4) Quarters Worked Upper Bound
Nonattainment$\times 1(\tau > 1971)$	199.71	418.30***	0.018**	0.030***
	(123.48)	(123.20)	(0.008)	(0.008)
Counties	148	148	148	148
First Year	1969	1969	1969	1969
Last Year	1974	1974	1974	1974
Sample Size	5376630	5374530	5376630	5374530
Weather Controls	Yes	Yes	Yes	Yes
Natality Basic	Yes	Yes	Yes	Yes
Natality Unrestricted	Yes	Yes	Yes	Yes

NOTE: This table reports regression coefficients from 4 separate regressions. The regressions are estimated using the individual-level micro data. The dependent variable in columns (1) and (2) is mean annual earnings between ages 29 and 31. The outcome variable in columns (3) and (4) is the annual mean quarters employed $\in [0, 4]$ from ages 29-31. The regression sample in column (1) and (3) drops the top 0.16 percent of individuals in the distribution of each post-CAAA cohort (to create a lower bound on the estimate). The regression sample in column (2) and (4) drops the bottom 0.16 percent of individuals in the distribution of each post-CAAA cohort (to create an upper bound on the estimate). The table reports estimates from reduced form models, estimating the effects nonattainment designation for cohorts born in the years after the CAAA went into effect. All regressions control for individual-level covariates (indicators for race, sex, and month of birth), a vector of contemporaneous state-level time-varying controls observed in the individual's state of work. Regressions also include birth-county fixed effects, birth-state$\times$birth-year fixed effects, and predetermined county characteristics (population, employment, total transfers) interacted with quadratic trends. Additional birth-county time-varying controls for weather and natality characteristics are described further in the text. Standard errors are clustered by commuting zone and are in parentheses. See text for details. Source: LEHD, EPA.

Table A17: The Effects of CAAA Implementation and TSP Exposure in the Year of Birth on Educational Attainment

	(1)	(2)	(3)	(4)
	Panel A: Nonattain and Education			
Nonattainment×1($\tau > 1971$)	0.029**	0.028**	0.025**	0.022*
	(0.014)	(0.013)	(0.012)	(0.012)
	Panel B: TSPs and Education			
Mean TSPs (/10)	-0.024*	-0.029**	-0.027**	-0.027*
	(0.013)	(0.014)	(0.013)	(0.016)
First Stage F-Statistic	14.7	15.0	14.8	14.2
Counties	148	148	148	148
First Year	1969	1969	1969	1969
Last Year	1974	1974	1974	1974
Sample Size	888	888	888	888
County FE	Yes	Yes	Yes	Yes
Weather Controls		Yes	Yes	Yes
Natality Basic			Yes	Yes
Natality Unrestricted				Yes

NOTE: This table reports regression coefficients from 8 separate regressions, 4 per panel. The dependent variable in all columns is educational attainment, as reported in the LEHD. The regression sample remains the same across all columns. Panel A reports estimates from reduced form models, estimating the effects nonattainment designation for cohorts born in the years after the CAAA went into effect. Panel B reports the 2SLS results from estimating the effects of the TSP level in the year of birth, using Nonattainment×1($\tau > 1971$) as an instrumental variable. Kleibergen-Paap Wald F statistics are reported as a test of the first-stage strength of the instrument. All regressions control for birth-county fixed effects, birth-state×birth-year fixed effects, and predetermined county characteristics (population, employment, total transfers) interacted with linear trends. Additional controls for weather and natality characteristics are listed in the table and described further in the text. The regressions are weighted by the birth-county×birth-year cell size. Standard errors are clustered by commuting zone and are in parentheses. See text for details. Source: LEHD, EPA.

Table A18: The Effect of CAAA Implementation on Public Assistance and Poverty in the 2000 Census and 2005-2012 ACS

	(1) Any Welfare	(2) Below Poverty	(3) Below 2×Poverty
Nonattainment×1($\tau > 1971$)	-0.001	-0.006**	-0.010***
	(0.001)	(0.003)	(0.003)
Counties	121	121	121
First Year	1969	1969	1969
Last Year	1974	1974	1974
Sample Size	696	696	696
Year of Birth Weather Controls	Yes	Yes	Yes
Year of Birth Natality Basic	Yes	Yes	Yes
Year of Birth Natality Unrestricted	Yes	Yes	Yes

NOTE: This table reports regression coefficients from 3 separate regressions. The dependent variable in each regression is a composition-adjusted labor market average for a birth-county×birth-year cell. The outcome in column (1) is the rate of welfare income receipt in a birth-county×birth-year; the outcome in column (2) is the proportion of individuals living in households with incomes below the poverty line in a birth-county×birth-year; the outcome in column (3) is the proportion of individuals living in households with incomes below twice the poverty line in a birth-county×birth-year. The regression sample remains the same in all columns and is limited to individuals born in 1969-1974 in one of the 24 LEHD sample states. Birth counties are assigned based on individuals' counties of residence in the survey year. The table reports estimates from reduced form models, estimating the effects nonattainment designation for cohorts born in the years after the CAAA went into effect. All regressions control for birth-county fixed effects, birth-state×birth-year fixed effects, and predetermined county characteristics (population, employment, total transfers) interacted with quadratic trends. Additional birth-county time-varying controls are listed in the table. The regressions are weighted by the sum of the survey weights in each birth-county×birth-year cell. Standard errors are clustered by commuting zone and are in parentheses. See text for details. Source: IPUMS, EPA.

Table A19: The Effect of CAAA Implementation on Worker Outcomes in the March 1987-1997 CPS

	(1) Earnings	(2) Unemployment	(3) Disability
Nonattainment×1($\tau > 1971$)	393.141	9.471	-18.562*
	(479.078)	(27.979)	(10.868)
Mean Dep. Var	8260	63.2	12.1
Counties	128	128	128
First Year	1969	1969	1969
Last Year	1974	1974	1974
Sample Size	11180	11180	11180

NOTE: This table reports regression coefficients from 3 separate regressions using the LEHD-CPS linked data on individuals surveyed in the 1987-1997 waves of the March CPS. The regressions are estimated using the individual-level micro data. The dependent variables are: self-reported annual earnings, the self-reported amount of income received from unemployment insurance, and the self-reported amount of income received from disability insurance. The regression sample remains the same in all columns and is limited to individuals born in 1969-1974 in one of the 24 LEHD sample states. The table reports estimates from reduced form models, estimating the effects nonattainment designation for cohorts born in the years after the CAAA went into effect. All regressions control for birth-county fixed effects, birth-state×birth-year fixed effects, and predetermined county characteristics (population, employment, total transfers) interacted with quadratic trends. Additional birth-county time-varying controls are listed in the table and described further in the text. The regressions are unweighted. Standard errors are clustered by commuting zone and are in parentheses. See text for details. Source: LEHD, CPS, EPA.

B Additional Results

In this Appendix, we present some further analyses that explore additional treatment effect heterogeneity as well as the robustness of our baseline estimates to using different imputation metrics for nonattainment and to using different samples of individuals from the LEHD. Lastly, we present estimates exploring whether contemporaneous exposure to air pollution is observably different across treatment and control cohorts.

Non-linear Effects of Air Pollution

The models in the main text assume that the effect of exposure to TSPs in one's year of birth is linear in exposure. However, it is possible that the dose-response relationship between air pollution and health is non-linear (Currie and Neidell, 2005; Schlenker and Walker, 2011). We explore the possible non-linear effects in two separate ways. First, we simply include a higher order polynomial in TSP in an OLS fixed effects framework. While we think the regression coefficient *magnitudes* may be biased by unobserved omitted factors correlated with both changes in TSP and long-run outcomes, the *shape* of the dose-response curve is likely less sensitive to these concerns (unless of course the bias varied with the level of pollution - which might happen through avoidance behavior such as "bad air day" alerts). We use this logic to explore the shape of the dose-response function by fitting OLS, fixed-effect regression models that include a higher order polynomial in TSP (i.e. a quadratic). Formally, we augment equation (2) with a quadratic term in the average annual county TSP level. Panel A of Appendix Table B1 presents the results. The results are generally not well estimated and mixed.

If the bias in the dose-response function is non-linear, then the OLS fixed effects framework will still produce biased estimates. We therefore use the CAAA quasi-experimental design to shed some light on possible non-linear impacts of pollution. Panel B of Table B1 presents models that of from our baseline reduced form specifications that also add a three-way interaction term between the county nonattainment indicator, the post-1971 birth cohort indicator, and the 1971 (pre-treatment) county pollution level: $Non_{1970,c} \times 1[\tau > 1971] \times TSP_{c,1971}$. This allows the dose-response curve to vary (linearly) in the 1971 level of TSP in a county. If this interaction term is zero, this would be consistent with the hypothesis that the marginal effect of a one unit increase in pollution is the same regardless of the level of ambient air pollution (i.e. a constant, linear dose-response). If the coefficient on the interaction was significantly positive, then this would be consistent with the hypothesis that the marginal effect of ambient air pollution on later life outcomes is progressively worse in areas with higher than average pollution levels. A challenge with this particular test is that the average level of ambient air pollution in a county can be correlated with many observed and unobserved factors that may contribute to heterogeneity in the dose-response relationship. For example, people in more polluted areas may lack basic preventive health services and thus be *more* responsive to marginal changes in air pollution because of their underlying health conditions rather than any non-linearity in the dose-response. The results in Table B1 provide some suggestive, but ultimately inconclusive, evidence of non-linearity. An alternative explanation for the linear response may be driven by the relatively small amount of variation we have in TSP changes in the data.

Alternative Imputations of Nonattainment Status

As noted in the text, we do not observe nonattainment status in our data. Thus far, we have followed the prior work in this literature (e.g., Chay and Greenstone (2003a) and Sanders and Stoecker (2015) among others), by defining nonattainment status based on EPA regulatory thresholds and 1970 pollution monitor readings in each county. We investigate the sensitivity of this imputation mechanism

in several ways. First, we use pollution data from other years to classify a county as nonattainment based on EPA mandated TSP thresholds. Appendix Table B2 shows the results from estimating our baseline reduced form and 2SLS regressions using different years to impute nonattainment status (1969 through 1971, with the 1970 column replicating our baseline results). The reduced form results across these specifications are very similar. However, the 2SLS models are less precise and only statistically significant when 1970 is used for imputation of nonattainment.[51] While previous researchers have used similar timing assumptions (e.g.: Chay and Greenstone, 2003a), we note this as a limitation.

As an alternative way of assigning nonattainment status, we use EPA data from 1978 and 1979, which are the first years to contain information both on pollution monitor readings and actual EPA-recorded nonattainment status. We run an auxiliary probit regression, where we regress nonattainment status in a county-year on the annual county-year TSP level interacted with the two thresholds. We use the estimates from this regression to form a predicted probability of nonattainment based on 1969, 1970, and 1971 county variables. We use this predicted probability of nonattainment in our baseline regression models in place of the binary nonattainment measure from before. Appendix Table B3 presents results, which are similar to the main ones we have presented.[52]

Including More States in the Analysis Sample

As we write in Section 4, our analysis focuses on the 24 LEHD states, which continuously contain earnings records during 1998-2007. We have explored robustness to removing the sample constraint, which requires that people be born in one of the 24 states AND work in one of the 24 states in our sample. Appendix Tables B4 and B5 present results where we let individuals work in any of the 30 states in the LEHD rather than the 24 states in our baseline sample. Results remain similar to our baseline estimates.

Different Ages of Follow-Up

We have explored the sensitivity of our results to measuring outcomes across different ages in several ways. First, we replicated our baseline analysis, controlling for age fixed effects in our first-step auxiliary regression (i.e., when controlling for other micro-level covariates such as race, sex, and month of birth). Tables B6 and B7 present results, which are nearly identical to the baseline results from before. We have also estimated models where we average labor market outcomes within an individual over a wider age bracket (i.e., from 28 to 32), while also including age fixed effects. Tables B8 and B9 present results. As before, the results remain similar to our baseline estimates.

[51] The first stage is weaker when we use the 1969 and 1971 years for imputation; hence it is to be expected that the less precisely estimated first stages from the other years yield less precise 2SLS estimates.

[52] Note that the standard errors in Panel A of Appendix Table B3 do not reflect the fact that nonattainment is a generated regressor. In standard OLS regression, inference using generated regressors should be corrected for auxiliary regression sampling variance (e.g. Murphy and Topel (2002)). However, when the generated regressor is used as an instrumental variable this need not be the case. Wooldridge (2002, p. 117) presents a weak set of assumptions for which the standard errors of 2SLS regressions using generated instruments are unbiased. The key assumption turns on strict exogeneity between the error term in the structural model and the covariates used to generate the instrument in the auxiliary regression. See Dahl and Lochner (2012) for a similar approach, using a predicted variable as an instrumental variable in a 2SLS setting. These issues are also discussed tangentially in Wooldridge (1997, 2003).

Results Without Winsorized Earnings

As described in Section 4 in the manuscript, we cap earnings at age 28 equivalent $100,000 allowing for 2% annual growth in earnings in order to limit the influence of outliers. Specifically, we cap earnings at $100,000 for 28 year olds, $102,000 for 29 year olds, $104,040 for 30 year olds, $106,121 for 31 year olds, and $108,243 for 32 year olds. Appendix Table B10 explores the sensitivity of our baseline estimates to removing the earnings cap. The results are qualitatively similar to our main findings but with slightly larger coefficients, suggesting that CAAA led to an increase in annual earnings of $228 to $357 in affected cohorts.

Table B1: Exploring the Functional Form of the Dose-Response Function Relating TSP Exposure in the Year of Birth to Earnings at Ages 29-31

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Panel A: Quadratic TSP, OLS							
	Quarters Worked				Earnings			
Mean TSPs (/10)	0.000	0.000	0.001	0.001	-12.81	-13.59	-5.86	-9.86
	(0.001)	(0.001)	(0.001)	(0.001)	(27.88)	(19.42)	(17.83)	(17.71)
Mean TSPs (/10)-Squared	0.000	0.000	0.000	0.000	0.358	0.378*	0.270	0.308
	(0.000)	(0.000)	(0.000)	(0.000)	(0.267)	(0.222)	(0.202)	(0.197)
	Panel B: 1971 TSP Interaction							
	Quarters Worked				Earnings			
Nonattain×1($\tau > 1971$)	0.020	0.026	0.035	0.034	101.31	51.80	81.30	-20.04
	(0.055)	(0.041)	(0.035)	(0.033)	(500.02)	(676.05)	(577.33)	(529.06)
Nonattain×1($\tau > 1971$)×TSP_{1971}	0.001	0.001	0.003	0.002	88.81	37.79	28.56	49.20
	(0.008)	(0.006)	(0.005)	(0.005)	(127.18)	(106.38)	(91.55)	(83.62)
Sample Size	888	888	888	888	888	888	888	888
County FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Weather Controls		Yes	Yes	Yes		Yes	Yes	Yes
Natality Basic			Yes	Yes			Yes	Yes
Natality Unrestricted				Yes				Yes

NOTE: This table reports regression coefficients from 16 separate regressions, 8 per panel. The dependent variable in each regression is a composition-adjusted labor market average for a birth-county×birth-year cell. The outcome variable in columns (1)-(4) is the annual mean quarters employed $\in [0, 4]$ from ages 29-31. The outcome variable in columns (5)-(8) is mean earnings from ages 29-31. The regression sample remains the same across all columns and panels. Panel A reports fixed effects, OLS estimates pertaining to the effect of ambient TSP levels (and a quadratic term in TSP) in the year of birth on labor market outcomes from the ages of 29-31. Panel B reports estimates from reduced form models, estimating the effects nonattainment designation for cohorts born in the years after the CAAA went into effect. The regression models in Panel B also include an interaction between the ambient level of TSP in 1971 and county nonattainment designation, as a test of treatment effect heterogeneity. All regressions control for cohort (i.e. birth-county) fixed effects, state×year fixed effects, and predetermined county characteristics (population, employment, total tranfers) interacted with quadratic trends. Additional controls for weather and natality characteristics are listed in the table and described further in the text. Regressions are weighted by the cohort size. Standard errors are clustered by commuting zone and are in parentheses. See text for details. Source: LEHD, EPA.

Table B2: The Effects of CAAA Implementation and TSP Exposure in the Year of Birth on Labor Market Outcomes at Ages 29-31, Different Nonattainment Imputations

	(1)	(2)	(3)	(4)	(5)	(6)
	Panel A: Nonattainment and Labor Market Outcomes					
	Quarters Worked			Earnings		
	1969	1970	1971	1969	1970	1971
Nonattainment$\times 1(\tau > 1971)$	0.020***	0.022***	0.025***	204.888*	276.665**	311.015***
	(0.008)	(0.008)	(0.007)	(118.200)	(115.459)	(105.026)
	Panel B: TSPs and Labor Market Outcomes					
	Quarters Worked			Earnings		
	1969	1970	1971	1969	1970	1971
Mean TSPs (/10)	-0.018	-0.028*	-0.019	-178.863	-351.745*	-237.997
	(0.012)	(0.016)	(0.012)	(143.240)	(190.392)	(158.050)
Sample Size	888	888	888	888	888	888
First Stage F-Statistic	9.38	14.2	3.37	9.38	14.2	3.37

NOTE: This table reports regression coefficients from 12 separate regressions, 6 per panel. The dependent variable in each regression is a composition-adjusted labor market average for a birth-county×birth-year cell. The outcome in columns (1)-(3) is the mean annual quarters employed, while the outcome in columns (4)-(6) is mean annual earnings. Each column reports results where county nonattainment status is assigned based on ambient TSP levels in each of the three years listed in the column headings (1969, 1970, and 1971). The regression sample remains the same across all columns and panels. Panel A reports estimates from reduced form models, estimating the effects nonattainment designation for cohorts born in the years after the CAAA went into effect. Panel B reports the 2SLS results from estimating the effects of the TSP level in the year of birth, using Nonattainment$\times 1(\tau > 1971)$ as an instrumental variable. Kleibergen-Paap Wald F statistics are reported as a test of the first-stage strength of the instrument. All regressions control for birth-county fixed effects, birth-state×birth-year fixed effects, and predetermined county characteristics (population, employment, total transfers) interacted with quadratic trends. Additional birth-county time-varying controls for weather and natality characteristics are described further in the text. The regressions are weighted by the birth-county×birth-year cell size. Standard errors are clustered by commuting zone and are in parentheses. See text for details. Source: LEHD, EPA.

Table B3: The Effects of CAAA Implementation and TSP Exposure in the Year of Birth on Labor Market Outcomes at Ages 29-31, Auxiliary Regression Nonattainment Imputation

	(1)	(2)	(3)	(4)	(5)	(6)
	Panel A: Nonattainment and Labor Market Outcomes					
	Quarters Worked			Earnings		
	1969	1970	1971	1969	1970	1971
Nonattainment$\times 1(\tau > 1971)$	0.037***	0.039***	0.047***	363.016**	455.163***	596.521***
	(0.011)	(0.012)	(0.010)	(175.187)	(174.466)	(161.661)
Sample Size	888	888	888	888	888	888
	Panel B: TSPs and Labor Market Outcomes					
	Quarters Worked			Earnings		
	1969	1970	1971	1969	1970	1971
Mean TSPs (/10)	-0.021*	-0.039*	-0.022*	-205.536	-454.089*	-282.620
	(0.012)	(0.021)	(0.013)	(141.782)	(246.196)	(172.088)
Sample Size	888	888	888	888	888	888
First Stage F-Statistic	13	14.9	3.86	13	14.9	3.86

NOTE: This table reports regression coefficients from 12 separate regressions, 6 per panel. The dependent variable in each regression is a composition-adjusted labor market average for a birth-county×birth-year cell. The outcome variable in columns (1)-(3) is the mean annual quarters employed, while the outcome variable in columns (4)-(6) is mean annual earnings. To assign nonattainment status, we use EPA data from 1978 and 1979, which contain information both on pollution monitor readings and actual nonattainment status as recorded by the EPA. We run an auxiliary probit regression, where we regress nonattainment status in a county-year on the two TSP variables used in nonattainment designation (i.e., the geometric mean and the 2nd highest daily TSP reading), and the interactions between these variables and the relevant thresholds. We use the estimates from this auxiliary regression to form a predicted probability of nonattainment based on 1969, 1970, and 1971 county pollution data. We use this predicted probability of nonattainment in our baseline regression models in place of the binary nonattainment measure from before. The regression sample remains the same across all columns and panels. Panel A reports estimates from reduced form models, estimating the effects nonattainment designation for cohorts born in the years after the CAAA went into effect. Panel B reports the 2SLS results from estimating the effects of the TSP level in the year of birth, using Nonattainment$\times 1(\tau > 1971)$ as an instrumental variable. Kleibergen-Paap Wald F statistics are reported as a test of the first-stage strength of the instrument. All regressions control for birth-county fixed effects, birth-state×birth-year fixed effects, and predetermined county characteristics (population, employment, total transfers) interacted with quadratic trends. Additional birth-county time-varying controls for weather and natality characteristics are described further in the text. The regressions are weighted by the birth-county×birth-year cell size. Standard errors are clustered by commuting zone and are in parentheses. See text for details. Source: LEHD, EPA.

Table B4: The Effect of CAAA Implementation on Labor Market Outcomes at Ages 29-31, Born in Sample States, Work in Any Available State

	(1)	(2)	(3)	(4)
	Panel A: Quarters Employed			
Nonattainment×1($\tau > 1971$)	0.022**	0.022**	0.021***	0.024***
	(0.010)	(0.009)	(0.008)	(0.007)
	Panel B: Annual Earnings			
Nonattainment×1($\tau > 1971$)	236.414*	270.934**	268.514**	288.401**
	(142.904)	(126.587)	(114.832)	(112.719)
Counties	148	148	148	148
First Year	1969	1969	1969	1969
Last Year	1974	1974	1974	1974
Sample Size	888	888	888	888
Year of Birth Weather Controls		Yes	Yes	Yes
Year of Birth Natality Basic			Yes	Yes
Year of Birth Natality Unrestricted				Yes

NOTE: This table reports regression coefficients from 8 separate regressions, 4 per panel using all 30 states with data available in the LEHD. The sample uses everyone who is in our baseline sample, but we allow them to work in any of the available LEHD states, rather than just the 24 states in our sample. The dependent variable in each regression is a composition-adjusted labor market average for a birth-county×birth-year cell. The outcome variable in Panel A is the mean annual quarters employed, and the outcome variable in Panel B is mean annual earnings. Both panels report estimates from reduced form models, estimating the effects nonattainment designation for cohorts born in the years after the CAAA went into effect. All regressions control for birth-county fixed effects, birth-state×birth-year fixed effects, and predetermined county characteristics (population, employment, total transfers) interacted with quadratic trends. Additional birth-county time-varying controls are listed in the table. The regressions are weighted by the birth-county×birth-year cell size. Standard errors are clustered by commuting zone and are in parentheses. See text for details. Source: LEHD, EPA.

Table B5: The Effect of TSP Exposure in the Year of Birth on Labor Market Outcomes at Ages 29-31, Born in Sample States, Work in Any Available State

	(1)	(2)	(3)	(4)
	Panel A: Quarters Employed			
Mean TSPs (/10)	-0.018*	-0.023*	-0.023**	-0.030*
	(0.011)	(0.012)	(0.011)	(0.016)
	Panel B: Annual Earnings			
Mean TSPs (/10)	-196.171	-279.344*	-286.358*	-366.987*
	(137.503)	(158.719)	(146.255)	(190.351)
First Stage F-Statistic	14.7	15	14.8	14.3
Counties	148	148	148	148
First Year	1969	1969	1969	1969
Last Year	1974	1974	1974	1974
Sample Size	888	888	888	888
Year of Birth Weather Controls		Yes	Yes	Yes
Year of Birth Natality Basic			Yes	Yes
Year of Birth Natality Unrestricted				Yes

NOTE: This table reports regression coefficients from 8 separate regressions, 4 per panel using all 30 states with data available in the LEHD. The sample uses everyone who is in our baseline sample, but we allow them to work in any of the available LEHD states, rather than just the 24 states in our sample. The dependent variable in each regression is a composition-adjusted labor market average for a birth-county×birth-year cell. The outcome variable in Panel A is the mean annual quarters employed, and the outcome variable in Panel B is mean annual earnings. The regression models estimate the effects of the TSP level in the year of birth, using Nonattainment$\times 1(\tau > 1971)$ as an instrumental variable. Kleibergen-Paap Wald F statistics are reported as a test of the first-stage strength of the instrument. The regressions control for birth-county fixed effects, birth-state×birth-year fixed effects, and predetermined county characteristics (population, employment, total transfers) interacted with quadratic trends. Additional birth-county time-varying controls are listed in the table. The regressions are weighted by the birth-county×birth-year cell size. Standard errors are clustered by commuting zone and are in parentheses. See text for details. Source: LEHD, EPA.

Table B6: The Effect of CAAA Implementation on Labor Market Outcomes at Ages 29-31, with Age Fixed Effects

	(1)	(2)	(3)	(4)
	Panel A: Quarters Employed			
Nonattainment×1($\tau > 1971$)	0.020**	0.023***	0.019**	0.022***
	(0.010)	(0.008)	(0.008)	(0.008)
Sample Size	888	888	888	888
	Panel B: Annual Earnings			
Nonattainment×1($\tau > 1971$)	224.60	278.75**	250.58**	276.67**
	(148.36)	(126.24)	(118.64)	(115.46)
Counties	148	148	148	148
First Year	1969	1969	1969	1969
Last Year	1974	1974	1974	1974
Sample Size	888	888	888	888
Year of Birth Weather Controls		Yes	Yes	Yes
Year of Birth Natality Basic			Yes	Yes
Year of Birth Natality Unrestricted				Yes

NOTE: This table reports regression coefficients from 8 separate regressions, 4 per panel. The dependent variable in each regression is a composition-adjusted labor market average for a birth-county×birth-year cell. The outcome variable in Panel A is the mean annual quarters employed, and the outcome variable in Panel B is mean annual earnings. In these specifications, we also control for age fixed effects in the first-step, auxiliary regression. Both panels report estimates from reduced form models, estimating the effects nonattainment designation for cohorts born in the years after the CAAA went into effect. All regressions control for birth-county fixed effects, birth-state×birth-year fixed effects, and predetermined county characteristics (population, employment, total transfers) interacted with quadratic trends. Additional birth-county time-varying controls are listed in the table and described further in the text. The regressions are weighted by the birth-county×birth-year cell size. Standard errors are clustered by commuting zone and are in parentheses. See text for details. Source: LEHD, EPA.

Table B7: The Effect of TSP Exposure in the Year of Birth on Labor Market Outcomes at Ages 29-31, with Age Fixed Effects

	(1)	(2)	(3)	(4)
	Panel A: Quarters Employed			
Mean TSPs (/10)	-0.017*	-0.024**	-0.021*	-0.028*
	(0.010)	(0.012)	(0.011)	(0.016)
	Panel B: Annual Earnings			
Mean TSPs (/10)	-186.21	-286.84*	-266.89*	-351.75*
	(139.30)	(158.61)	(145.62)	(190.39)
First Stage F-Statistic	14.7	15	14.8	14.2
Counties	148	148	148	148
First Year	1969	1969	1969	1969
Last Year	1974	1974	1974	1974
Sample Size	888	888	888	888
Year of Birth Weather Controls		Yes	Yes	Yes
Year of Birth Natality Basic			Yes	Yes
Year of Birth Natality Unrestricted				Yes

NOTE: This table reports regression coefficients from 8 separate regressions, 4 per panel. The dependent variable in each regression is a composition-adjusted labor market average for a birth-county×birth-year cell. The outcome variabe in Panel A is the mean annual quarters employed, and the outcome variable in Panel B is mean annual earnings. In these specifications, we also control for age fixed effects in the first-step, auxiliary regression. Both panels report the 2SLS results from estimating the effects of the TSP level in the year of birth, using Nonattainment$\times 1(\tau > 1971)$ as an instrumental variable. Kleibergen-Paap Wald F statistics are reported as a test of the first-stage strength of the instrument. All regressions control for birth-county fixed effects, birth-state×birth-year fixed effects, and predetermined county characteristics (population, employment, total transfers) interacted with quadratic trends. Additional birth-county time-varying controls are listed in the table. The regressions are weighted by the birth-county×birth-year cell size. Standard errors are clustered by commuting zone and are in parentheses. See text for details. Source: LEHD, EPA.

Table B8: The Effect of CAAA Implementation on Labor Market Outcomes at Ages 27-35, with Age Fixed Effects

	(1)	(2)	(3)	(4)
	Panel A: Quarters Employed			
Nonattainment×1($\tau > 1971$)	0.017*	0.019**	0.017*	0.018**
	(0.009)	(0.009)	(0.009)	(0.009)
Sample Size	888	888	888	888
	Panel B: Annual Earnings			
Nonattainment×1($\tau > 1971$)	176.75	245.02**	229.33*	255.92**
	(126.25)	(122.51)	(117.61)	(116.39)
Counties	148	148	148	148
First Year	1969	1969	1969	1969
Last Year	1974	1974	1974	1974
Sample Size	888	888	888	888
Year of Birth Weather Controls		Yes	Yes	Yes
Year of Birth Natality Basic			Yes	Yes
Year of Birth Natality Unrestricted				Yes

NOTE: This table reports regression coefficients from 8 separate regressions, 4 per panel. The dependent variable in each regression is a composition-adjusted labor market average for a birth-county×birth-year cell. The outcome variable in Panel A is the mean annual quarters employed, and the outcome variable in Panel B is mean annual earnings. In these specifications, we also control for age fixed effects in the first-step, auxiliary regression. Both panels report estimates from reduced form models, estimating the effects nonattainment designation for cohorts born in the years after the CAAA went into effect. All regressions control for birth-county fixed effects, birth-state×birth-year fixed effects, and predetermined county characteristics (population, employment, total transfers) interacted with quadratic trends. Additional birth-county time-varying controls are listed in the table and described further in the text. The regressions are weighted by the birth-county×birth-year cell size. Standard errors are clustered by commuting zone and are in parentheses. See text for details. Source: LEHD, EPA.

Table B9: The Effect of TSP Exposure in the Year of Birth on Labor Market Outcomes at Ages 27-35, with Age Fixed Effects

	(1)	(2)	(3)	(4)
	Panel A: Quarters Employed			
Mean TSPs (/10)	0.015	0.017*	0.017*	0.020**
	(0.011)	(0.011)	(0.011)	(0.011)
	Panel B: Annual Earnings			
Mean TSPs (/10)	-209.05	-258.95*	-253.15**	280.22**
	(153.31)	(149.46)	(140.74)	(142.73)
First Stage F-Statistic	15.5	15.9	15.7	15.1
Counties	148	148	148	148
First Year	1969	1969	1969	1969
Last Year	1974	1974	1974	1974
Sample Size	888	888	888	888
Year of Birth Weather Controls		Yes	Yes	Yes
Year of Birth Natality Basic			Yes	Yes
Year of Birth Natality Unrestricted				Yes

NOTE: This table reports regression coefficients from 8 separate regressions, 4 per panel. The dependent variable in each regression is a composition-adjusted labor market average for a birth-county×birth-year cell. The outcome variable in Panel A is the mean annual quarters employed, and the outcome variable in Panel B is mean annual earnings. In these specifications, we also control for age fixed effects in the first-step, auxiliary regression. Both panels report the 2SLS results from estimating the effects of the TSP level in the year of birth, using Nonattainment$\times 1(\tau > 1971)$ as an instrumental variable. Kleibergen-Paap Wald F statistics are reported as a test of the first-stage strength of the instrument. All regressions control for birth-county fixed effects, birth-state×birth-year fixed effects, and predetermined county characteristics (population, employment, total transfers) interacted with quadratic trends. Additional birth-county time-varying controls are listed in the table. The regressions are weighted by the birth-county×birth-year cell size. Standard errors are clustered by commuting zone and are in parentheses. See text for details. Source: LEHD, EPA.

Table B10: The Effects of CAAA Implementation and TSP Exposure in the Year of Birth on Earnings at Ages 29-31, non-Winsorized Earnings

	(1)	(2)	(3)	(4)
	Panel A: Nonattainment and Earnings			
Nonattainment×1($\tau > 1971$)	228.473	357.356**	312.351*	336.038*
	(187.260)	(169.653)	(161.095)	(175.476)
	Panel B: TSP Pollution and Earnings			
Mean TSPs (/10)	-189.579	-367.890*	-332.742*	-427.723*
	(158.966)	(191.887)	(175.900)	(232.164)
First Stage F-Statistic	14.8	15.1	14.9	14.3
Counties	148	148	148	148
First Year	1969	1969	1969	1969
Last Year	1974	1974	1974	1974
Sample Size	888	888	888	888
County FE	Yes	Yes	Yes	Yes
Weather Controls		Yes	Yes	Yes
Natality Basic			Yes	Yes
Natality Unrestricted				Yes

NOTE: This table reports regression coefficients from 8 separate regressions, 4 per panel. The dependent variable in each regression is a composition-adjusted labor market average for a birth-county×birth-year cell. The outcome variable is mean annual earnings (not winsorized at $100,000). The regression sample remains the same across all columns and panels. Panel A reports estimates from reduced form models, estimating the effects nonattainment designation for cohorts born in the years after the CAAA went into effect. Panel B reports the 2SLS results from estimating the effects of the TSP level in the year of birth, using Nonattainment×1($\tau > 1971$) as an instrumental variable. Kleibergen-Paap Wald F statistics are reported as a test of the first-stage strength of the instrument. All regressions control for birth-county fixed effects, birth-state×birth-year fixed effects, and predetermined county characteristics (population, employment, total transfers) interacted with quadratic trends. Additional birth-county time-varying controls for weather and natality characteristics are described further in the text. The regressions are weighted by the birth-county×birth-year cell size. Standard errors are clustered by commuting zone and are in parentheses. See text for details. Source: LEHD, EPA.

C Data Appendix

C.1 Matching Algorithm

In order to match Clean Air Act nonattainment status and county pollution levels at birth to individuals' earnings records, we take advantage of the LEHD variables that detail each worker's place of birth. Derived from Social Security Administration records, the two variables "pobcity" (place of birth) and "pobst" (state, or foreign born, country of birth) should in theory identify the county of birth. However, in practice, a major challenge with the data is that they are string variables (filled out by hand when applying for a Social Security card) that can be difficult to match to County FIP codes; while the place of birth variable is supposed to list the city of birth (or county, if born in a non-incorporated community), there are almost 2,280,002 unique combinations of Pobcity and Pobst among the non-foreign born in the data whereas there are fewer than 22,800 actual Census places and counties in the United States, implying over two million errors in the string variables. The several orders of magnitude more "places of birth" in the LEHD stem from (using the example of Los Angeles) misspellings (e.g. Las Angeles), alternative abbreviations (e.g. L.A.), extraneous information (e.g. Los Angeles, Los Angeles County), neighborhoods within the city (e.g. Hollywood), incorrect states (e.g. CO), some combination thereof, and other errors. In order to match these string variables to county FIP codes in both an accurate and comprehensive manner, we follow a 5-step process. In each step, we draw on the Census Bureau's database of Census places, counties, and minor civil divisions as well as the United States Geological Survey's Geographic Names Information System (GNIS) file, which includes information on non-Census geographic localities.

Step 1: We take the LEHD string variables as is and merge them to the string variables from the Census and GNIS databases. The process successfully matches up 68,698 unique combinations of pobcity and pobst from the LEHD to their county fips codes, which constitutes 3.01% of the total unique combinations of the two variables, and when weighted by the number of workers in the LEHD that report each combination, 56.12% of the sample. Broken down by file type, 16,956 (0.75% unweighted and 48.86% weighted) are matched to Census places, 2,063 to counties (0.09% unweighted and 1.41% weighted), 2,585 to minor civil divisions (0.12% unweighted, 0.61% weighted), and 47,094 to GNIS localities (3.01% unweighted and 5.24% weighted).

Step 2: We split up the pobcity string variable into separate strings delimited by spaces (to split up the string variable into separate words) and then merge the first word from each string to the string variables from the Census and GNIS databases. We carry out this step because examination of several thousand string variables by hand indicated that the most prominent "error" in the data occurs when the city is listed followed by one or more letters that represent the county in which the city is located. In this step, we make one additional restriction; the first word from the pobcity variable cannot be matched when that word also matches the first word of a two or more word geographic locality (i.e. if both "Los" and "Los Angeles" exist, we do not match up a string with a first word of "Los"). The process successfully matches up 528,126 unique observations, representing 23.16% of all unique combinations unweighted and 29.48% of weighted combinations. Broken down by file type, 329,650 are matched to Census places (14.43%% unweighted and 25.38% weighted), 29,837 to counties (1.34% unweighted and 0.87% weighted), 17,432 to minor civil divisions (0.76% weighted and 0.48% unweighted), and 151,207 to GNIS localities (6.63% unweighted and 2.75% weighted).

Step 3: We split up the pobcity string variable into separate strings delimited by spaces and then merge twenty different combinations of the multiple strings variables to the string variables from the Census and GNIS databases (e.g. the first two words of the string, the first three words of the string,

etc). A full list of these combinations is available upon request. This process successfully merges 15,846 "places of birth", which constitutes 0.68% of all listed places of birth (0.92% weighted)

Step 4: We split the pobcity string variable up into separate strings delimited by spaces and replace the first word in the string if they appear to abbreviate a direction (e.g. N, No, Nor, S, etc) or the word Saint (St, Snt) with the full word (e.g. North, Saint), and then merge the new string variable to the string variables from the Census and GNIS databases. The process successfully merges 7,003 unique combinations, representing 0.32% of all unique combinations unweighted and 0.28% of weighted combinations.

Step 5: We take the remaining combinations (1,660,729, 72.82% unweighted and 13.20% weighted) and select high frequency combinations (defined as those for which at least 100 individuals in the data report that combination or that shares the first word in pobcity with a combination that at least 100 individuals report). We then take these combinations (98,705, 4.33% unweighted, and 10.77% weighted) and attempt to match them by hand. This process is primarily undertaken to deal with high frequency spelling mistakes. We are able to hand match 82,928 of these combinations, representing 3.64% of all unique combinations unweighted and 10.32% of weighted combinations.

Duplicates: There are several sources of duplicate localities within a state that share the same name but are in different counties. Whenever a Census place or county shares a name with a minor civil division or non-Census locality, we assume that the individual was born in either the Census place or county (since the proper way to fill out the form is to list either the city or county of birth). If there are multiple cities within a state in different counties or multiple minor civil divisions or non-Census localities within the same state in different counties, we treat these observations as unmatched (80,256 combinations, 3.52% unweighted and 1.31% weighted). On the other hand, if a city and county share the same name but the city is in a different county, we assume the correct place of birth is the city (as the large majority of individuals list their city), unless the string indicates that it is a county (e.g. the last word in the string is "Co", "Cou", "Count", "County", etc). However, the results are robust to making the opposite assumption or treating these observations as unmatched. The following table shows the distribution of matches by step. In all, we match 622,345 unique combinations of pobst and pobcity, which constitutes 27.30% of all unique LEHD combinations and 95.81% of LEHD individuals' places of birth.

Table C1: Match Statistics

	Number of unique places of birth	Percent of total	Percent of total (weighted by no. of persons)	Cumulative Weighted Match Rate
Step 1	68,698	3.01	56.12	56.12
Step 2	528,126	23.16	29.48	85.60
Step 3	15,846	0.68	0.92	86.52
Step 4	7,003	0.32	0.28	86.80
Step 5	82,928	3.64	10.32	97.12
Duplicates	80,256	-3.52	-1.31	95.81
Total match	622,345	27.30	95.81	
Total unmatched	1,657,657	72.70	4.19	

NOTE: Source: LEHD, USGS

Table C2: Cohort Age Coverage in the LEHD

Age	Birth Years	Earnings Years
28	1970-1979	1998-2007
29	1969-1978	1998-2007
30	1968-1977	1998-2007
31	1967-1976	1998-2007
32	1966-1975	1998-2007

C.2 Creating Composition Adjusted Earnings Measures

We use a composition adjusted labor market measure in most all of the specifications in the paper. To create this composition adjusted earnings measure, we use a two-step method that controls for micro-determinants of earnings heterogeneity before collapsing the data to the birth-county×birth-year level. Composition adjusted earnings measures are common in both the labor and urban economics literatures. See e.g., Angrist and Lavy (2009); Baker and Fortin (2001); Currie *et al.* (2015); Donald and Lang (2007); Albouy (2009a,c); Notowidigdo (2011); Shapiro (2006), among others.

Specifically, the estimating equations in the two-step method are:

$$\text{Step 1: } y^a_{ict\tilde{s}\tilde{t}} = W'_{ict}\pi + Z'_{\tilde{s}\tilde{t}}\lambda + \theta_{ct} + \epsilon_{ict\tilde{s}\tilde{t}} \tag{7a}$$

$$\text{Step 2: } \widehat{\theta_{ct}} = \phi_0 + \phi_1 TSP_{ct} + X'_{ct}\rho + \gamma_c + \eta_{st} + \nu_{ct} \tag{7b}$$

Equation (7a) is first estimated using the individual-level micro data. In terms of notation, any subscript with a tilde measures a contemporaneous relationship (i.e., $\tilde{t}$ represents the year the worker was observed working at age a, while $\tilde{s}$ represents the individual's state of work), while subscripts without a tilde refer to information in the year of birth. The outcome, $y^a_{ict\tilde{s}\tilde{t}}$, is annual earnings or quarters employed for an individual i of age a born in county c and year t, observed working in year $\tilde{t}$ and state $\tilde{s}$. The explanatory variables in equation (7a) include: W_{ict}, which is a vector of all available individual-level covariates (indicators for race, sex, and month of birth), as well as $Z_{\tilde{s}\tilde{t}}$, which is a vector of contemporaneous state-level time-varying controls.[53] Equation (7a) also controls for a full set of birth-county×birth-year indicators, θ_{ct}. This regression yields coefficient estimates, $\widehat{\theta_{ct}}$, which are the covariate-adjusted average labor market outcomes for a birth-county×birth-year, after controlling for race, sex, month of birth, and the work-state×year characteristics.

The second stage equation (7b) uses the conditional mean outcomes for a birth-county×birth-year ($\widehat{\theta_{ct}}$) from equation (7a) as the dependent variable. Note that equation (7b) is the exact same regression model as equation (2), except the dependent variable is now $\widehat{\theta_{ct}}$. In other words, we use the group-level, composition-adjusted averages rather than the unadjusted group means as outcomes. Similar to equation (2), equation (7b) includes controls for a vector of time-varying socio-economic, demographic, and climate characteristics in the county and year of birth that may also influence earnings determination and labor force participation at ages 29-31 (X_{ct}); county fixed effects that control for time-invariant, unobserved determinants of labor market outcomes for workers born in a particular county (γ_c); and birth-state×birth-year fixed effects that control for time-varying determinants of long-run outcomes that are common to all individuals born in a particular state×year (η_{st}).

[53] The exact work-state level controls (all in per-capita terms) are: Supplemental Nutrition Assistance Program (SNAP) spending, Medical benefit spending, public medical care benefits, family assistance spending, personal transfer receipts, income maintenance, and unemployment insurance benefits. The state-level variables are constructed by taking population weighted averages over all counties in a state×year from the REIS data.