NBER WORKING PAPER SERIES

BUILDING THE PROTOTYPE CENSUS ENVIRONMENTAL IMPACTS FRAME

John L. Voorheis
Jonathan M. Colmer
Kendall A. Houghton
Eva Lyubich
Mary Munro
Cameron Scalera
Jennifer R. Withrow

Working Paper 31189
http://www.nber.org/papers/w31189

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
April 2023

Any opinions and conclusions expressed herein are those of the author(s) and do not reflect the views of the U.S. Census Bureau or the National Bureau of Economic Research. The authors thank Surya Menon and Yolande Tra for research support, and to Nick Muller, Corbett Grainger and the participants at the AERE Annual Meeting, CRIW Conference on Natural Capital Accounting and Census Bureau research seminars for helpful comments. The Census Bureau has reviewed this data product to ensure appropriate access, use, and disclosure avoidance protection of the confidential source data used to produce this product (Data Management System (DMS) number: P-7505723, Disclosure Review Board (DRB) approval numbers: CBDRB-FY2022-CES010-017, CBDRB-FY23-CES014-003, CBDRB-FY23-CES014-005, CBDRB-FY2023-CES010-004 and CBDRB-FY23-CES019-009)

Building the Prototype Census Environmental Impacts Frame
John L. Voorheis, Jonathan M. Colmer, Kendall A. Houghton, Eva Lyubich, Mary Munro, Cameron Scalera, and Jennifer R. Withrow
NBER Working Paper No. 31189
April 2023
JEL No. Q53,Q54

ABSTRACT

The natural environment is central to all aspects of life, but efforts to quantify its influence have been hindered by data availability and measurement constraints. To mitigate some of these challenges, we introduce a new prototype of a microdata infrastructure: the Census Environmental Impacts Frame (EIF). The EIF provides detailed individual-level information on demographics, economic characteristics, and address-level histories – linked to spatially and temporally resolved estimates of environmental conditions for each individual – for almost every resident in the United States over the past two decades. This linked microdata infrastructure provides a unique platform for advancing our understanding about the distribution of environmental amenities and hazards, when, how, and why exposures have evolved over time, and the consequences of environmental inequality and changing environmental conditions. We describe the construction of the EIF, explore issues of coverage and data quality, document patterns and trends in individual exposure to two correlated but distinct air pollutants as an application of the EIF, and discuss implications and opportunities for future research.

John L. Voorheis
US Census Bureau
john.l.voorheis@census.gov

Jonathan M. Colmer
Department of Economics
University of Virginia
j.colmer@virginia.edu

Kendall A. Houghton
US Census Bureau
kendall.a.houghton@census.gov

Eva Lyubich
US Census Bureau
eva.lyubich@census.gov

Mary Munro
MITRE
mmunro@mitre.org

Cameron Scalera
US Census Bureau
cameron.j.scalera@census.gov

Jennifer R. Withrow
US Census Bureau
jennifer.withrow@census.gov

1 Introduction

The air we breathe, the water we drink, the food we eat, and the climate we inhabit all inextricably shape economic activity and human flourishing. In addition to providing essential inputs for production, environmental goods and services directly affect our health and our ability to learn, our capacity to work, our productivity when we do work, and many other dimensions of well-being.

In recent decades, we have fundamentally affected the integrity of our biosphere, our climate, our oceans, our freshwater systems, and our land systems (Secretariat of the Convention on Biological Diversity, 2020; IPCC, 2021; IPBES Secretariat, 2021). Economists and ecologists have long theorized that reductions in natural capital will result in large, and potentially irreversible, social costs (Frank and Schlenker, 2016; IPCC, 2022; Dasgupta, 2021), especially when there is limited substitutability between natural and human-made capital (Arrow and Fisher, 1974; Dasgupta and Heal, 1974; Stiglitz, 1974; Solow, 1993; Brock and Xepapadeas, 2003; Weitzman, 2009). However, despite growing interest, efforts to assess and understand how economic activity and the environment influence one another have been plagued by data availability and measurement constraints (Heal, 2000; Fenichel and Abbott, 2014; Ferraro et al., 2019). Further complicating analysis, environmental benefits (or damages) are unlikely to be evenly distributed across individuals within a population, making inference from aggregated data sources difficult.

To help mitigate some of these challenges, we introduce the prototype of a new microdata infrastructure to facilitate individual-level analyses of environmental conditions, their causes, and their consequences, in the United States – the Census Environmental Impacts Frame (EIF). The EIF uses confidential Census Bureau microdata drawn from surveys, administrative records, and Decennial censuses to provide detailed panel data on demographic and economic characteristics and address-level residential histories for nearly all residents of the United States from the late 1990s forward. This rich microdata (containing more than 6 billion observations) presents new opportunities to advance our understanding of the en-

vironmental conditions people face, why differences in exposure to environmental conditions arise, and the distributional consequences of exposure in the United States.

While advances in technology have improved our understanding of *where* environmental amenities and hazards are located, there remain large gaps in our understand about *who* is exposed, as well as the causes and consequences of these exposures. Existing work studying environmental quality frequently uses place-based data combining spatial data on the environment with neighborhood-level demographic characteristics such as population shares for different race and ethnic groups and median income. This can result in misleading inferences about individual level exposure relative to group exposure – a problem known as the ecological fallacy. Where individual-level data has been used it has either been limited by smaller samples in survey data, affecting statistical power, or it has lacked information on demographic characteristics, such as in administrative tax records. This has prevented serious study of differences in exposure across individual demographic characteristics. The EIF is not constrained along these dimensions. With precise individual-level data, ecological fallacies and aggregation bias can be avoided. With detailed demographic and economic data for the near-population of the United States, researchers using the EIF can engage seriously with questions related to heterogeneity.

The use of comprehensive individual-level data unlocks at least four new avenues of inquiry that are not feasible when using place-based data:

1. Using the exact address of individuals allows for a more precise characterization of the distribution of environmental quality and its consequences across individuals within a population. This is particularly important for minimizing the risk of ecological fallacies and aggregation bias when exposures vary substantially within even narrowly defined geographies such as Census tract.[1]

2. When documenting differences in exposure to environmental quality, individual-level

[1]One potential high value statistical production use of the EIF will be to inform and improve the measurement of community vulnerability and resilience, for instance in extending the US Census Bureau's Community Resilience Estimates to measure resilience to climate impacts.

data facilitates a deeper, intersectional understanding of exposure. Instead of documenting differences between places based on single demographic dimensions such as the median income of a location, or the share of the population that is of a particular demographic group, we are able to document differences in exposures along multiple dimensions, e.g. differences in exposure within a given income group, by race, for homeowners vs. renters, etc.

3. Detailed residential histories create an opportunity to characterize how exposure to environmental quality varies over the life cycle, construct cumulative exposures that account for migration, and evaluate how migration contributes to evolving patterns and trends.

4. The individual level data facilitates opportunities to explore the degree to which the consequences of, or responses to, environmental exposures vary between individuals that experience the same exposure, i.e., holding constant the level of environment exposure we can evaluate how differences in the characteristics of individuals determine outcomes.

The EIF marks an important shift in the data frontier — a single, well-curated microdata infrastructure that allows quantitative research on environmental quality to move from a place or community-based accounting to one that centers around individuals and their intersectional identities. The EIF will be updated as new data becomes available, allowing us to extend the economic panel and incorporate advances in the measurement of environmental quality as computational advances and new frontier measures become available. When finalized, we plan to make the EIF available to qualified researchers on approved projects through the Federal Statistical Research Data Centers (FSRDCs), a network of 31 secure physical locations. Researchers must apply and have projects approved before access. The FSRDCs provide a proven, secure mode of data distribution that can safeguard the privacy and confidentiality of the extensive microdata contained in the Environmental Impacts

Frame. Access to the EIF through the FSRDC network will provide extensive opportunities for researchers to conduct research that was previously not possible.

The remainder of this paper presents an overview of the data and an application to illustrate its capabilities, focused on how the EIF can facilitate deeper consideration of issues around Environmental Justice and the distribution of environmental hazards. Section 2 describe the different data sets used as inputs to construct the prototype EIF. Section 2 provides details on the build process. Section 3 provides details on population coverage and data quality. In Section 4, we present facts from ongoing research using the EIF about how the distribution of fine particulate matter ($PM_{2.5}$) and nitrogen oxide (NO_X) air pollution varies between different race, ethnicity, and income groups, and how these patterns have evolved over time. Finally, in Section 4.2 we illustrate how the lessons learned from using individual-level data can contribute to a distributional approach to natural capital accounting. Section 5 concludes.

2 Building the Prototype Environmental Impacts Frame

The Environmental Impacts Frame is a modular infrastructure, rather than a single combined file. There are two core modules constructed from confidential data at the Census Bureau: a demographic "spine", which contains basic information for the relevant universe of individuals, and an annual residential history file (RHF) for individuals in the spine. In this section, we discuss the construction of each of these modules.

These files can be combined to form one large panel dataset, but in practice they are maintained as separate modules. Analyses can proceed by using the residential histories combined with the spine for universe level analysis, or by combining some other relevant data (such as survey data from the American Community Survey) with either the spine or residential histories. This modular setup maximizes researcher flexibility, and provides a framework for subsequent expansions of scope (for instance, future work on housing charac-

teristics, employment histories, and family structure will be incorporated as separate linkable modules).

Constructing the Demographic Spine

The foundation of the EIF is the Census Bureau Numident – an administrative records dataset of individuals who have applied for Social Security Numbers (SSNs). Individuals in the Numident are assigned a Protected Identification Key (PIK), a consistent identifier across records within the Census data infrastructure, which allows researchers to link the same individual across different data sets.[2]

We use the Census Numident to construct a list of linkable individuals in the EIF. We refer to this list as the "spine", which form a backbone upon which we can merge all other information.[3] We select all PIKs from the most recent Census Numident and maintain PIK, date of birth, place of birth, the indicator whether the individual was born outside the United States, date of death (if relevant), and sex. We attach race information using an internal Census Bureau file that aggregates racial information from various survey and administrative records called the Title 13 best race and ethnicity file.

We clean place-of-birth locations and assign place of birth state and county FIPS codes to individuals, using the November 2021 version of the place of birth geographic crosswalk (created as part of the Decennial Census Digitization and Linkage Project). Only those with a birthplace in one of the 50 states or Washington, DC are included in the place of birth crosswalk, as individuals born overseas or in the US territories do not have detailed place of birth information in the spine. Of those born in one of the 50 states or Washington, DC, we are able to add a state and county FIPS to over 98% of PIKs. The EIF maintains those born elsewhere without cleaned place-of-birth information.

We use date of birth information directly from the Numident. We produce a composite

[2]See Wagner and Layne (2014) for more details on the Census Bureau's linkage process.

[3]We borrow the "spine" terminology from similar efforts to construct longitudinal administrative records data sets, e.g. Davis-Kean et al. (2017).

death date variable by first drawing from the Numident, then from Medicare data if this information is not in the Numident. In the rare case that there is still no information, we use information on death from the VSGI data.

Constructing the Residential History File

We work to create a residential history for each person in the Numident. The goal of this exercise is to select just one address per individual per year. We do this by searching for a given PIK across multiple administrative records that include address information, and selecting the single address that is likely to be the most accurate. Below, we describe each administrative record used and its initial cleaning, followed by details of the data build.

Administrative Tax Data

Administrative records from the Internal Revenue Service (IRS) constitute our preferred source of address information. Specifically, we use data from IRS 1040s (Individual Tax Forms) – including two special extracts of 1040s, the Electronic Filing (ELF) form, and the Modernized e-File (MeF) – and IRS 1099s (information returns).

The 1040 data contains the self-reported filer address at the time of filing.[4] These data have been delivered to Census annually from tax year 1998 to present. Within the 1040 data, individuals are listed as primary, secondary, or dependent filers. The 1040s may list the same individual in multiple different entries. Therefore, we select a single entry per individual based on the following criteria. If 1040s ever list an individual as a primary filer, we select that entry. If the form ever lists an individual as a secondary filer but never a primary filer, we select that entry. Lastly, if the only data for an individual is as a dependent,

[4] As a result, the income receiving period for the 1040s can deviate from the timing of location information (since the tax filing season is the calendar year following the tax year in which income was received). We will use processing year information for the residential histories, specifying cases where tax year information is used.

we select that entry.[5]

The paper version of the 1040 forms does not have space to list all the dependents in a given household, only up to four. Due to legacy processing constraints by the IRS, the main digitized 1040 data only retains these 4 dependents per return. We supplement these 1040 returns with additional information from returns electronically filed returns, which contain information on all dependents in a tax unit. These E-filed returns are processed in two different datasets provided by IRS, and can be linked to the main 1040 dataset by primary filer's PIK.

Lastly, we supplement address information with information returns, which are third party forms reporting payments (for instance, wages or interest) sent by the payer firm to both the IRS and individuals who received income in a tax year.[6] We have 1099 forms starting in tax year 2003 (processing year 2004). When using information return data, we select an address based on form type, prioritizing the form we believe to have the most accurate address information.[7] Our preference ordering over form types is W-2 >1098 >SSA-1099 >1099-R >1099-DIV >1099-INT >1099-S >1099-MISC >1099-G.

Secondary Data Sources

If we do not find an address for an individual in one of the tax-based sources, we use secondary data sources, including data from Medicare (Medicare EDB), the Department of Housing and Urban Development (HUD), the USPS National Change of Address (NCOA), the Veterans' Service Group of Illinois (VSGI), and a Census internal data source known as the Master Address File (MAF-ARF).

- We use administrative records from the Medicare Enrollment Database (EDB), which

[5]Due to a legacy processing error for some files delivered to the Census in the 2000s, we sometimes observe individuals listed as secondary tax filers in tax units with single filing status. In this case we categorize the individual listed as a secondary tax filer as a dependent instead.

[6]These information returns are referred to as 1099s colloquially, but include a variety of forms, including W-2s, 1098s and SSA-1099s.

[7]Businesses are more likely to have an updated address for their employees than, e.g. banks or other intermediaries may have for account holders, who have less frequent contact with payees.

contains addresses of enrollees. We observe these records from 1999-2021. For each individual, we select the address corresponding to their latest observation in a given year.

- The Master Address File Auxiliary Reference File (MAF-ARF) is a composite data set of individual and address pairs, which the Census Bureau prepares internally and releases annually. These PIK-MAFID pairs are assembled from all of the federally sourced administrative records held by the Census Bureau. We have MAF-ARF data covering the time period 2000-2021. The data set has a unique entry per individual per year.

- We use administrative records from two HUD programs: the Public and Indian Housing Certification program and the Tenant Rental Assistance Certification System. Data from these programs are harmonized and combined into a single longitudinally consistent file containing all individuals who are participants in HUD-assisted rental housing in a given year. We observe the address of the individual in the program as well as the effective date of their certification at that address. We observe these records from 1999-2021. For each individual, we select the address corresponding to their latest observation in a given year.

- The United States Postal Service National Change Of Address (NCOA) forms database provides address information when individuals submit a form to USPS requesting mail forwarding to a new address. We have records from 2010-2021. If we observe multiple entries for an individual in NCOA, we select the entry with the latest effective date. If there are multiple entries with the same effective date, we select an entry at random. Entries record both a "from" individual and a "to" individual for the mail forwarding service. We select the entry information from the "from" individual, unless this information is missing, in which case we select the entry from the "to" individual.

- The Veterans' Service Group of Illinois (VSGI) data is a third party dataset, which

the Census Bureau purchases. VSGI collects address information from the USPS, commercial vendors and other proprietary sources. VSGI data covers the time period 2015-2021. Each address has an effective date. For instances in which a person has multiple entries in a given year, we select the entry with the latest effective date.

Selecting a Single Location for Each Year

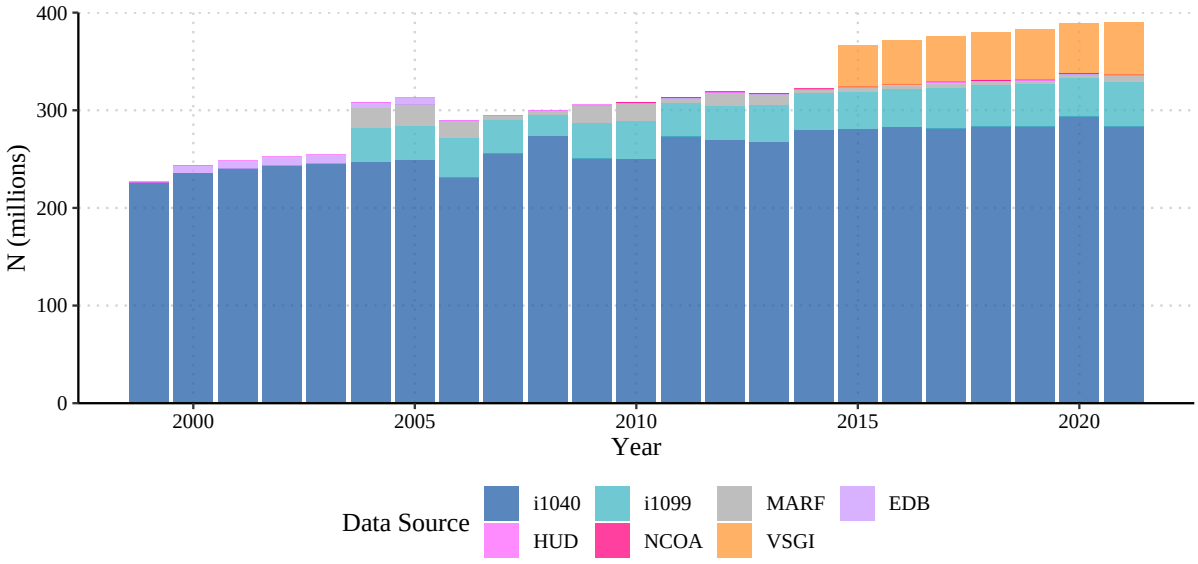
We select a single location for each individual in each year by first attempting to select a location entry from the most preferred source, proceeding to the second source if this entry is missing in the first source, and so on. Our preference ordering for data sources is: 1040, EDB, 1099, MAF-ARF, HUD, NCOA, VSGI. If the location variable is not present in any input dataset, we leave the entry as missing.

Location information consists of a MAFID and a ZIP code. MAFIDs are numeric IDs associated with the Master Address File (MAF) - a Census Bureau list of housing units in the United States. ZIP code information in our input data comes in two levels of detail - a five digit ZIP code only (zip5) and an additional 4 digits to the zip5 (zip9 or zip+4). We prioritize input data with zip+4 included. If zip5 and zip+4 are available, we select this information following our data source ranking.

We independently assign MAFID and ZIP codes based on the priority levels above. The ZIP codes and MAFIDs for a given PIK will not always correspond to the same source data due to the way we have chosen to prioritize different location information.

Finally, we construct the longitude and latitude of every individual's residential address in each year. We use the longitude and latitude of the MAFID when MAFID is present. When MAFID is not available, but ZIP code information is, we use the longitude and latitude of the ZIP code "centroid". For this purpose, the centroid is mean longitude and latitude of all people's residences within that ZIP code (the population-weighted center). Figure 1 shows the data source from which we construct the geographic coordinate locations of PIKs in the EIF. The majority of PIKs are assigned an address using 1040s.

Figure 1: Address Data Source



Source: Environmental Impacts Frame Spine and Residential History File, 1999-2021. **Notes:** See Section 2 for details on construction. Figure shows the source of address data for individuals in the EIF.

3 Evaluating Coverage and Data Quality

The EIF relies on the Census Bureau's data linkage infrastructure to combine multiple sources of administrative and demographic data to provide the most comprehensive information available on residential histories and basic demographics for United States residents. However, while these data are high quality, some sub-populations are not covered in the data and other sub-populations may be may be underrepresented.

By construction, we exclude all individuals who do not have a Social Security Number (SSN). This is because holding a SSN is a necessary condition for inclusion in the Census Numident. This restriction means that we do not have information on undocumented migrants and other residents who do not have a SSN. We are also likely to miss almost all individuals who do not have a connection to the formal economy in a given year. Two of our key input datasets – IRS 1040s and information returns – cover activity related to employment, asset

ownership, and other aspects of the formal economy. As such, individuals who are not employed or are otherwise disconnected from the formal economy may be systematically missing from these data. Similarly, data such as the Medicare EDB or the HUD administrative data only cover individuals who are eligible and take up these programs, while the USPS NCOA data relies on self-reported moves. The EIF combines these datasets to provide robustness to these selection issues, but it is not possible to rule out the possibility that some individuals will be missing for non-random reasons.

To investigate the potential for these issues, we first explore some basic descriptive facts about the EIF and do several data quality diagnostics. We then explore how the population in the EIF compares to the overall US population and to very large nationally representative samples of the US population using the American Community Survey (ACS). We conclude that the EIF covers a very large fraction of the US population, but that differences in coverage by race and ethnicity persist, although they have declined over time.

Table 1 presents summary statistics for two populations: 1) all those included in the Demographic spine and 2) those individuals in the spine who have information in the residential history files for 2005 and 2019. We present descriptive statistics for two years: 2005 and 2019. These years represent two different regimes in terms of data availability. 2005 includes the 1040s, 1099s, HUD, MAF-ARF, and Medicare EDB sources. The NCOA and VSGI data are only available starting in 2010 and 2015, respectively. The spine information represents all non-deceased individuals for these years, but the spine population may deviate from the US population for two reasons: 1) Numident death information is incomplete in earlier years, so many SSNs who died before the 1990s do not have a valid date of death (see (Finlay and Genadek, 2021)) and 2) individuals who received an SSN but no longer reside in the United States will appear as non-deceased in the spine. Therefore, while there are more people who appear non-deceased in the spine in 2019 than reside in the United States, there are only 323 million who have non-missing coordinates (approximately 98 percent of the 2019 population estimates.) As a result, in both 2005 and 2019, we see that the spine

population is slightly older than the EIF population. The EIF also has a slightly larger share of non-Hispanic White and Hispanic observations than the spine.[8]

Table 1: Sample Demographics

	2005		2019	
	Spine	Spine+RHF	Spine	Spine+RHF
Female	0.49	0.51	0.49	0.51
Mean Age	40.61	38.12	44.81	40.58
<18	0.24	0.23	0.20	0.20
65+	0.17	0.14	0.23	0.18
Hispanic	0.13	0.14	0.14	0.15
NH White	0.59	0.67	0.51	0.54
NH Black	0.12	0.12	0.11	0.11
NH Asian	0.04	0.04	0.03	0.04
Observations	369,500,000	277,500,000	390,800,000	323,900,000

Source: Environmental Impacts Frame Spine and Residential History File, 1999-2021. **Notes:** See Section 2 for details on construction. Table shows characteristics of non-deceased individuals in the spine and non-deceased individuals with an address on the residential history file.

3.1 Comparing the EIF with Other Populations

Beyond the basic descriptive statistics presented in Table 1, we are also interested in how well the EIF covers the overall US population and key sub-populations of interest. To do this comparison, we need to have a well defined US population to compare to. One possibility would be to compare total population counts in each year of the EIF residential history file to official population estimates from the US Census Bureau. However, this may not be an ideal comparison, as the official population estimates measure the total US

[8]This undercount in the spine is due to the fact that not all individuals can be assigned race/ethnicity information using the best race and ethnicity file, and these missing race Spine individuals appear to be disproportionately Hispanic. Future iterations of the EIF will incorporate composite files that include information from the 2020 Decennial Census, which should partially alleviate this.

resident population, including individuals categorically excluded from the EIF (most notably undocumented immigrants and other residents who do not have SSNs). Instead, we explore coverage in two ways: first, we consider what fraction of non-deceased PIKs in the EIF spine appear in the residential history file (and hence have geographic coordinates) in a given year, and second, we subset to a representative national sample (linkable cases in the American Community Survey). Each of these approaches have their own limitations – in the former case we are uncertain about the size of the emigrant and deceased population, and in the latter case there may be bias in PIK assignment or survey coverage error – but collectively they provide a more robust understanding of coverage.

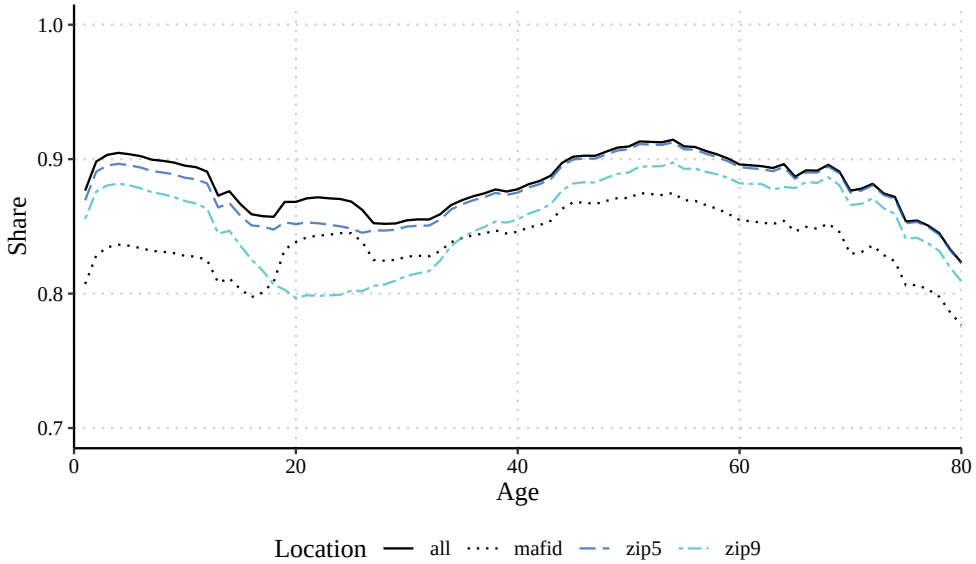
Overall, EIF coverage compared to the spine has increased over time, with around 80 percent of non-deceased PIKs appearing in the 2000 EIF residential history data, compared to around 90 percent in 2021. Note that the spine contains all individuals who applied for SSNs, including those who were but are no longer residents of the United States, as well as deceased individuals who do not have a date of death on the Numident. Finlay and Genadek (2021) presents evidence that the latter concern is particularly acute for individuals who died before 1970, while the former concern may be more acute during earlier years of the EIF. The number of nondeceased PIKs in the 2000 spine is 15% larger than the 2000 decennial census count, while the 2020 spine is only 5% larger than the 2020 Decennial count.[9]

We are also interested in understanding the extent to which there are differences in coverage between demographic groups. First, we consider coverage within a single EIF year (2015, the first year all our source datasets are available), focusing specifically on how this coverage varies over the lifecycle. Figure 2 shows coverage rates compared to the spine between the ages of one and 80. This coverage is highly non-linear, with declining coverage until around age 30, increasing coverage until around age 50, and then declining coverage afterwards. These rises and declines in coverage rates are likely due to age variation in the Numident from which the spine is derived – emigration is more likely at younger ages and

[9]See decennial Census and intercensal population estimates here: https://www.census.gov/programs-surveys/popest.html

unrecorded deaths are more likely for the over 50 population. Short run changes in coverage may also reflect variation in administrative data coverage. For example, spikes in coverage at age 18 and age 65 may reflect greater administrative data coverage as young adults register for selective service (a key source of the MAF-ARF data) and as older adults become eligible for and enroll in Medicare.

Figure 2: Share of Spine in 2015 EIF Residential History File, By Age



Source: Environmental Impacts Frame Spine and Residential History File, 2015. **Notes:** See Section 2 for details on construction. This figure shows the share of individuals in the spine also present in the 2015 EIF residential history file, by age. 2015 was chosen as it is the first year that all our source datasets are available.

Next, we turn to longitudinal considerations of coverage, focusing on coverage by gender, which is illustrated in Figure 3. Coverage has increased for both male and female PIKs on the spine, although in every year, we have slightly more coverage for female PIKs than male PIKs.

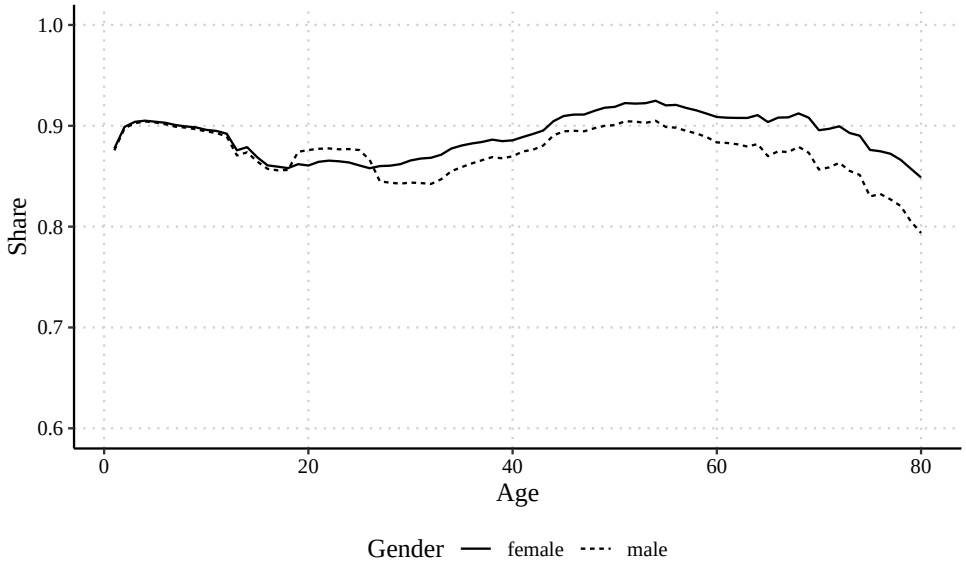
Figure 3: Share of Spine in EIF, By Gender (1999-2021)

Source: Environmental Impacts Frame Spine and Residential History File, 1999-2021. **Notes:** See Section 2 for details on construction. Figure shows the share of individuals in the spine also present in the 1999-2021 EIF residential history files, by gender.

Figure 4 explores these gender difference more by presenting coverage by gender in 2015 over the life cycle. We see that men and women have equal coverage up until adulthood, after which we see an increase in coverage for men, which we attribute to selective service registration. After this point women consistently have a greater share of coverage. This difference after early adulthood may be due to a greater propensity for men to be incarcerated[10] or a higher likelihood of being disconnected from the formal economy, both of which could result in worse coverage in the underlying administrative data compared to women. Men are also more likely to die at younger ages, so the under-coverage of deaths before the 1990s could mechanically impact the coverage rate for men.

[10]In future versions of the EIF we intend to incorporate more criminal justice data from the CJARS database, to explore this hypothesis.

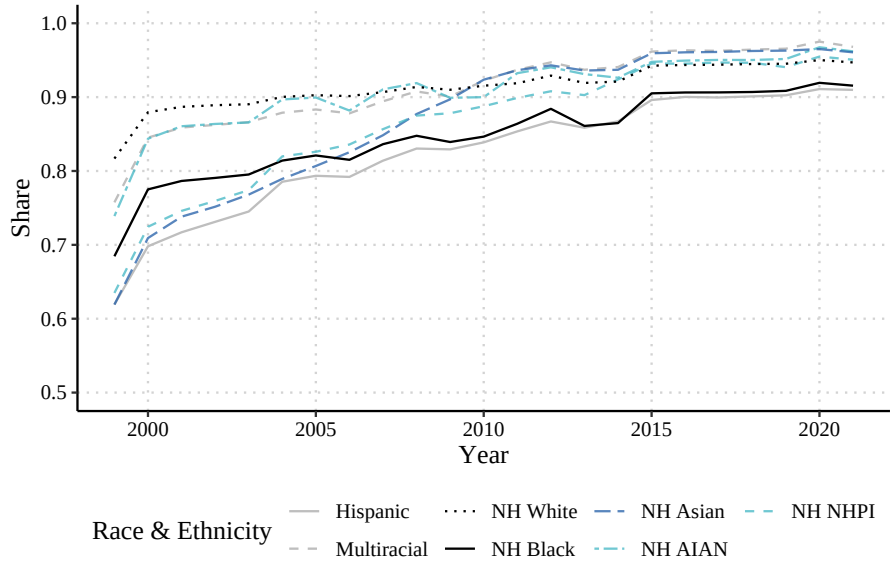
Figure 4: Share of Spine in 2015 EIF, By Age and Gender



Source: Environmental Impacts Frame Spine and Residential History File, 1999-2021. **Notes:** See Section 2 for details on construction. Figure shows the share of individuals in the spine also present in the 1999-2021 EIF residential history files, by age and gender.

Finally, we turn to coverage by race and income. We first document the trends in coverage relative to the spine by race and ethnicity in figure 5. We see here that there that trends in coverage have improved for all race and ethnic groups, but there remain large level gaps between groups: non-Hispanic Black and Hispanic individuals of any race have substantially lower coverage than non-Hispanic White individuals across all years of the EIF, although the Black-White and Hispanic-White coverage gaps have declined over time. Non-Hispanic Asian individuals have seen the largest improvements in coverage. In 1999, the Asian coverage rate was almost 20 percentage points lower than the non-Hispanic White coverage rate; by 2021, the Asian coverage rate was actually higher than the White coverage rate.

Figure 5: Share of Spine in EIF, by Race & Ethnicity



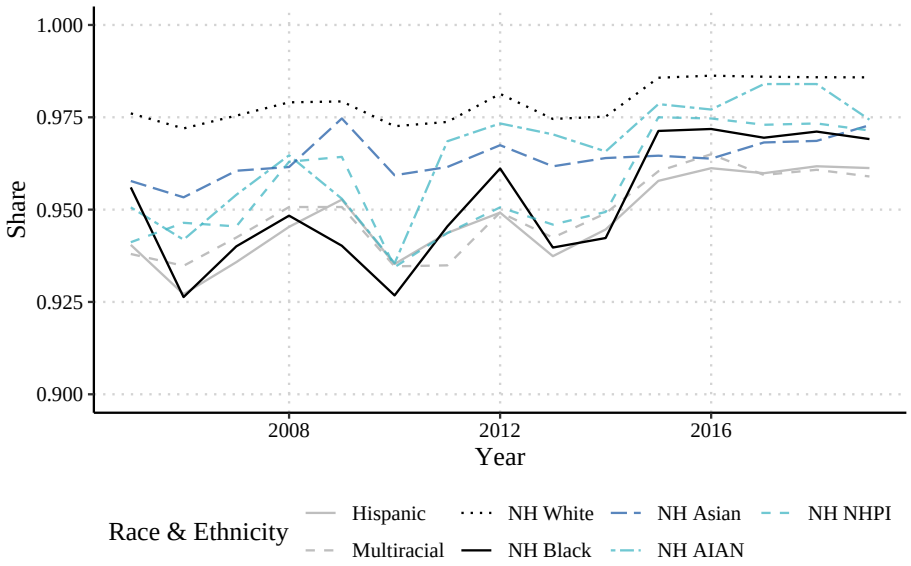
Source: Environmental Impacts Frame Spine and Residential History File, 1999-2021. **Notes:** See Section 2 for details on construction. Figure shows the share of individuals in the spine also present in the 1999-2021 EIF residential history files, by race and ethnicity.

As noted, analyzing coverage in relation to the spine presents some drawbacks in the form of uncertainty around emigration and unrecorded deaths. As an alternate approach we consider coverage relative to a known sample of individuals. We operationalize this using individuals who have responded to the American Community Survey (ACS) between 2005-2019. Unlike the spine, all individuals in the each annual ACS sample were alive at the time of their response, and importantly, were residing in the United States. Restricting to this sample, and using only ACS cases with PIKs as a denominator, results in substantially improved coverage rates (around 5-10 percentage points higher), with similar patterns of improvement over time as seen in the spine-based analysis.

In Figure 6 we examine how coverage has varied across race and ethnicity using the ACS sample as a benchmark. Consistent with the results using the spine as a benchmark, coverage is highest for non-Hispanic White individuals and lowest for Hispanic individuals of any race. However, coverage for non-Hispanic Black individuals is higher (relative to other

groups) using the ACS as a benchmark.

Figure 6: Share of ACS in EIF, by Race & Ethnicity

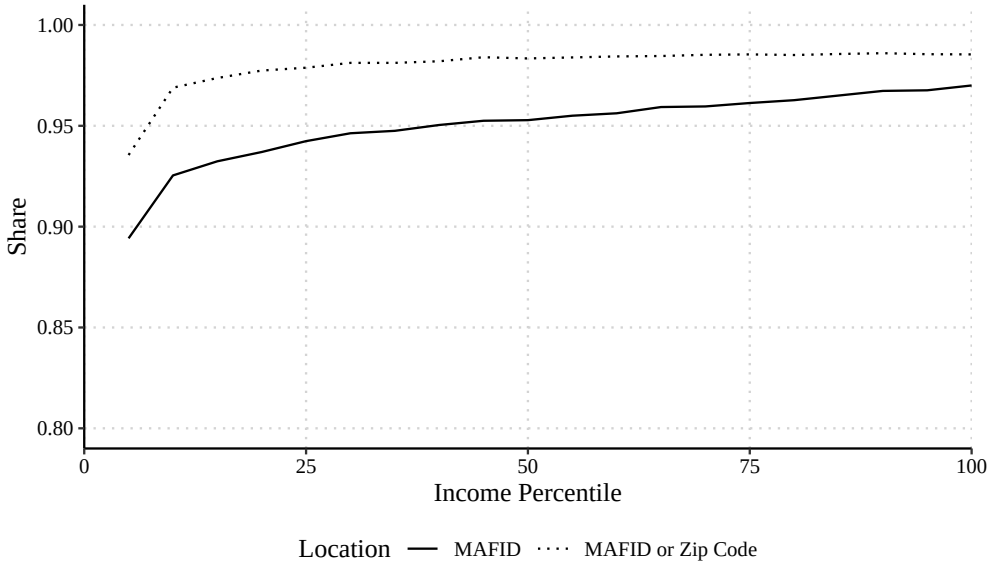


Source: Environmental Impacts Frame Residential History File, 2005-2019 and American Community Survey 2005-2019. **Notes:** See Section 2 for details on construction. Figure shows the share of individuals in the ACS also present in the 2005-2019 EIF residential history files, by race and ethnicity.

In Figure 7, we examine how coverage has varied across the 2019 ACS household income distribution.[11] We observe that lower income individuals have lower coverage rates than higher income individuals. We do, however, see that gaps in coverage between low and high income individuals have been shrinking over time, consistent with the decline in other coverage gaps.

[11]It would theoretically be possible to calculate coverage rates by income using administrative records in addition to the ACS, however, as noted above, the IRS 1040 records are a key component of the EIF residential histories, and as such any coverage rates relative to these data would be structurally biased towards 100 percent.

Figure 7: Share of ACS in 2019 EIF, by Income



Source: Environmental Impacts Frame Residential History File, 2019 and American Community Survey, 2019. **Notes:** See Section 2 for details on construction. Figure shows the share of individuals in the ACS also present in the 2019 EIF residential history files, by income.

Overall, coverage rates for the EIF are encouragingly high, suggesting that direct use of this infrastructure should capture a set of individuals close to the actual US population. Importantly, this coverage has improved over time, suggesting that cross-sectional analyses using the most recent data will be using the highest possible quality data. As noted, there remain small differences in coverage across groups which researchers should be aware of. In particular, coverage rates are slightly lower for men, for lower income individuals, and, depending on the year, for Hispanic and most non-White race groups. Future work is required to develop an appropriate strategy to adjust the EIF residential histories to be representative of the overall US population.

4 Incorporating Environmental Data

The EIF exists to facilitate the description and analysis of exposure to environmental amenities and hazards at the individual level. The core components described above – the demographic spine and the residential histories – provide the foundation for an exciting and extensive research agenda. Using the EIF, it is possible to incorporate and analyze any environmental data that can be geospatially resolved.

The EIF allows researchers to develop a systematic and comprehensive understanding of environmental exposures at a high spatial resolution over a relatively long time series. Environmental data that are currently being connected to the EIF are derived from administrative and remotely sensed data products, including: air pollution concentrations; wildfire burn perimeters, wildfire risk metrics, and wildfire smoke plume data; historical flood inundations, rainfall, and flood risk metrics; hurricane wind field exposure; surface temperature, air temperature, and heat wave measures; proximity to Superfund sites, polluting facilities, brownfields, and other fixed points of interest; projected sea level rise inundation; and airborne toxic releases. Multiple ongoing research projects use this data to provide comprehensive evidence on the distribution of exposures, their causes, and their consequences (Colmer et al., 2023a; Chakma et al., 2023; Burke et al., 2023; Colmer et al., 2023c,b).

To provide an illustration of how the EIF can be used, we present a case study that draws on past and ongoing work focused on understanding disparities in exposure to air pollution in the United State (Currie et al., 2020; Colmer et al., 2023a,c). First, we document patterns and trends in exposure to ambient air pollution, highlighting how individual exposures have evolved over time; second, we document how the distribution of air pollution exposure varies by race, by income, and by race within the income distribution. Drawing on these results, we present new evidence on how the distribution of exposures can be incorporated into a natural capital accounting framework.

4.1 Patterns and Trends in Individual-Level Air Pollution Exposure

In the last two decades, our understanding of how air pollution affects health and economic activity has dramatically expanded. It is now well established that even acute exposure to pollution can have immediate, persistent, and even intergenerational effects on a wide range of outcomes, including health, educational attainment, learning, decision-making, productivity, criminal activity, labor supply, and earnings (Chay and Greenstone, 2003; Currie and Neidell, 2005; Graff Zivin, J. and Neidell, M., 2012; Schlenker and Walker, 2015; Chang et al., 2016; Isen et al., 2017; Chang et al., 2018; Deryugina et al., 2019; Colmer and Voorheis, 2021; Colmer et al., 2023a). Alongside this expanding body of evidence, it is well established that disadvantaged communities are disproportionately exposed to higher levels of pollution (Commission for Racial Justice, United Church of Christ, 1987; Mohai et al., 2009; Banzhaf et al., 2019).

Much of the existing work on environmental justice has explicitly focused on proximity to polluting sites in communities, examining the degree to which neighborhood demographics are related to the presence of polluting sites (toxic waste facilities, industrial facilities, power plants, etc). The key fact that emerges from this body of work is that communities of color – specifically neighborhoods with large or predominant populations of Black and Hispanic people – and low income communities are much more likely to be close to polluting facilities than higher income or majority White communities. There is also evidence of large disparities in exposure to particulate matter by race (Colmer et al., 2020; Liu et al., 2021; Jbaily et al., 2022), and that these disparities have been mitigated in absolute terms by environmental policy (Currie et al., 2020). However, it remains the case that in a relative sense, environmental inequality has remained stubbornly persistent: the most polluted neighborhoods decades ago remain the most polluted neighborhoods today (Colmer et al., 2020).

This body of evidence convincingly documents that environmental amenities and hazards

vary across communities. We understand much less about how these exposures vary across individuals. The EIF's use of confidential microdata provides the framework to bridge this gap between people and places, which we illustrate by examining the distribution of exposure to air pollution.

We intersect the EIF with recently available satellite-derived data products (van Donkelaar et al., 2021; Cooper et al., 2022), which combine satellite observations, ground monitors, and statistical models to produce pollution concentration estimates at a very fine grid across the contiguous United States.[12] The data provide granular information on two major types of air pollution – $PM_{2.5}$ and Nitrogen Dioxide (NO_X). For each dataset, we intersect the gridded pollution data for each year with the corresponding residential history year file coordinates, assigning exposure to each individual based on the grid point in which their best latitude and longitude falls. We collect demographic characteristics for each individual from the spine. We estimate average annual exposure by race and ethnicity using a mutually exclusive categorization of race and ethnicity, focusing on the four largest groups: Hispanic of any race, Non-Hispanic White, Non-Hispanic Black, and Non-Hispanic Asian.[13] The $PM_{2.5}$ data from van Donkelaar et al. (2021) combines observations of Aerosol Optical Depth from satellite instruments with observations from ground level $PM_{2.5}$ monitors in a Geographic Weighted Regression model.[14] This allows for the prediction of ground level $PM_{2.5}$ level for grid cells without monitors. Similarly, the NO_X data from Cooper et al. (2022) combines satellite derived observations of the Vertical Column Depth (VCD) of Nitrogen Dioxide from two satellite instruments with ground level observations from NO_X pollution monitors to predict NO_X for grid cells without monitors. Using this data, we explore how trends in exposure have evolved for different race and income groups over the past two decades.

[12]The fine grid is 0.01 degrees, ∼1 km at the equator. Coverage of satellite data is incomplete at high latitudes, and one of the data products has a bounding box around North America, so we omit Alaska and Hawaii from subsequent pollution analysis.

[13]Results for Non-Hispanic American Indian and Alaska Native, and all other Non-Hispanic groups, including Native Hawaiian and Pacific Islanders, those with two or more races, and individuals, as well as those who cannot be assigned a race will be available in subsequent releases.

[14]Aerosol Optical Depth is a measure of visual occlusion within an image pixel sensed from a satellite such as MODIS.

Using the EIF provides several key advantages: 1) since the EIF represents a very large fraction of the overall population, it is possible to characterize exposure even for smaller groups which have seen less attention in the environmental justice literature due to sample size restrictions; 2) since residential histories are resolved to latitude and longitude coordinates, it is possible to assign exposure at residence instead of the average exposure within a county, Census tract, or Census block group (commonly used when working with public-use data); 3) since the EIF uses composite administrative records to measure race and ethnicity (which are consistently defined), it is straightforward to measure exposure differences between groups in a consistent way.[15] The data offers improvements not only over approaches which use aggregated demographic data to study disparities (Colmer et al., 2020), but also over previous microdata approaches which relied entirely on survey data (Currie et al., 2020).

Figure 8 presents $PM_{2.5}$ exposure for these four main race and ethnicity groups over time between 1999 and 2021. Panel a) reports the levels of exposure for each group in each year. Panel b) reports gaps relative to non-Hispanic White exposure in each year. Consistent with Currie et al. (2020), we document substantial improvements in air quality for all groups since the late 1990s. We also document substantial reductions in absolute disparities between non-Hispanic Black and non-Hispanic White individuals – the Black-White $PM_{2.5}$ gap narrowed from 1.5 $\mu g/m^3$ in 1999 to less than 0.5 $\mu g/m^3$ in 2020. This narrowing continues the trend documented in (Currie et al., 2020), who reported a decline in absolute disparities between 2000 and 2016. We note, however, that reductions in absolute disparities have not been universal for all groups. While Hispanic-White and Asian-White $PM_{2.5}$ gaps narrowed between 1999 and 2010 they have since returned to or even exceeded the gaps documented in 1999. One possible explanation for this increase in environmental inequality is the increasing severity of wildfire smoke, which disproportionately affects regions of the United States with large Hispanic and Asian populations. This hypothesis is explored in more detail by Burke

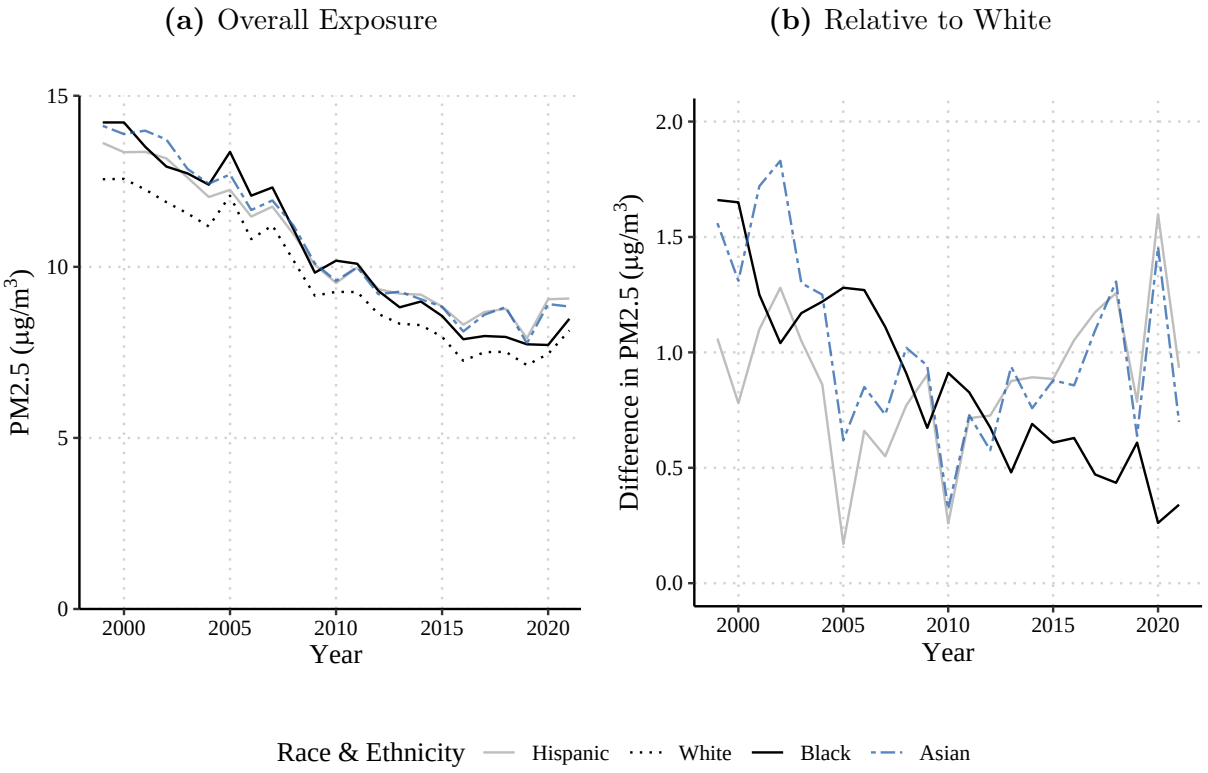
[15]For instance, Currie et al. (2020) focuses solely on Black-White differences, since these racial categories are more consistently defined across decennial Census and ACS data than other groups such as those reporting Hispanic ethnicity.

et al. (2023), who use new estimates of wildfire smoke $PM_{2.5}$ in combination with the EIF to explore the contribution of wildfire smoke to pollution disparities in the United States.

Figure 9 presents NO_X exposure for each of the reported race and ethnicity groups between 2005 and 2019.[16] Similar to $PM_2.5$, we observe a sharp decline in exposure for all groups over this period (Panel a). There remain significant absolute disparities in exposure, all groups have NO_X concentrations that are higher than non-Hispanic White individuals. While these gaps have declined, concentrations in 2019 for other groups are similar to or continue to exceed the concentrations Non-Hispanic White individuals were exposed to in 2005. Unlike $PM_{2.5}$ we do not see any evidence of a reversal in trends in either overall exposure absolute disparities. This is likely due to differences in sources between the two pollutants. Almost all NO_X emissions come from diesel exhaust or coal, and the period since 2005 has seen marked declines in coal use in power generation. By contrast, $PM_{2.5}$ comes from a multitude of primary and secondary sources. While the trends of increasing coal-to-gas fuel switching (and state and federal regulation of NO_X emissions) would be expected to influence both NO_X and $PM_{2.5}$ ambient concentrations, there are no countervailing influences (such as wildfire smoke) for NO_X.

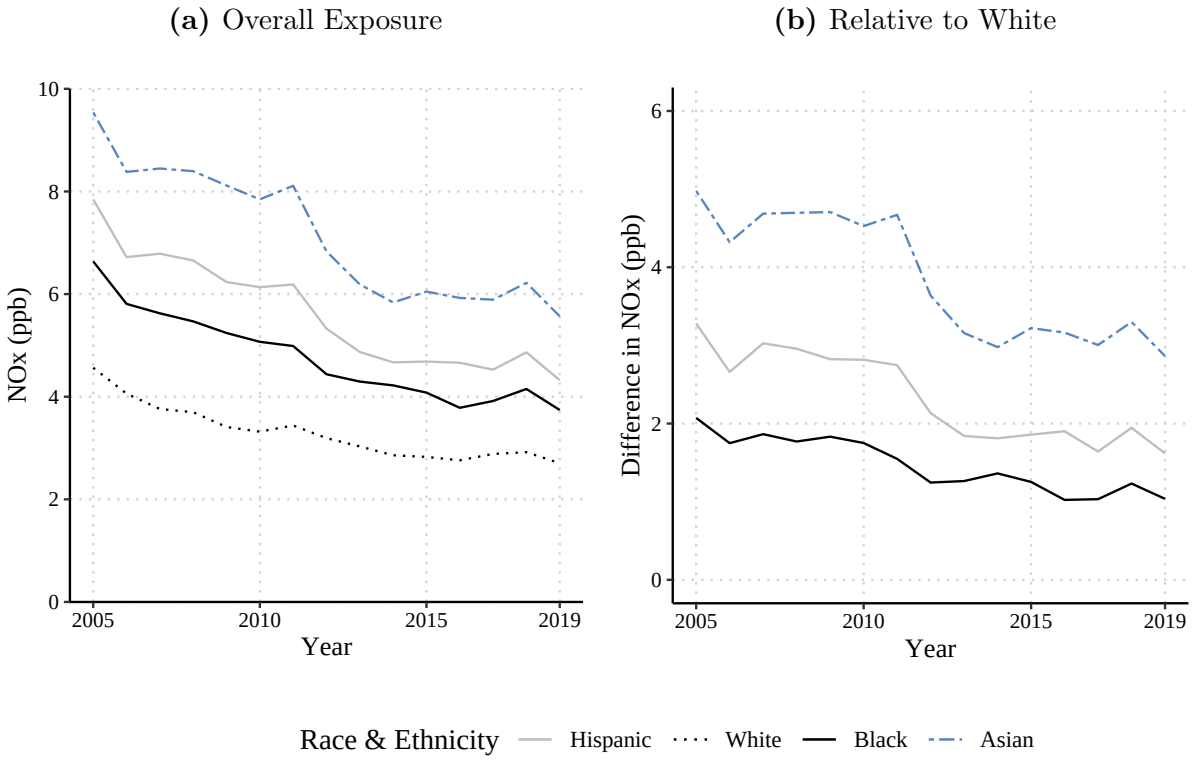
[16] $PM_{2.5}$ and NO_X data are available over slightly different time periods based on when the relevant satellites were launched — high resolution AOD data was first captured in 1998 by the MODIS satellite, while high resolution VCDs were not captured until the OMI satellite was launched, starting in 2005.

Figure 8: Racial and Ethnic Disparities in $PM_{2.5}$ Exposure



Source: Environmental Impacts Frame Residential History File, 1999-2021, and van Donkelaar et al. (2021). **Notes:** See Section 2 for details on construction. Figure shows trends in Nitrogen Dioxide Exposure by Race and Ethnicity.

Figure 9: Racial and Ethnic Disparities in NO_X Exposure



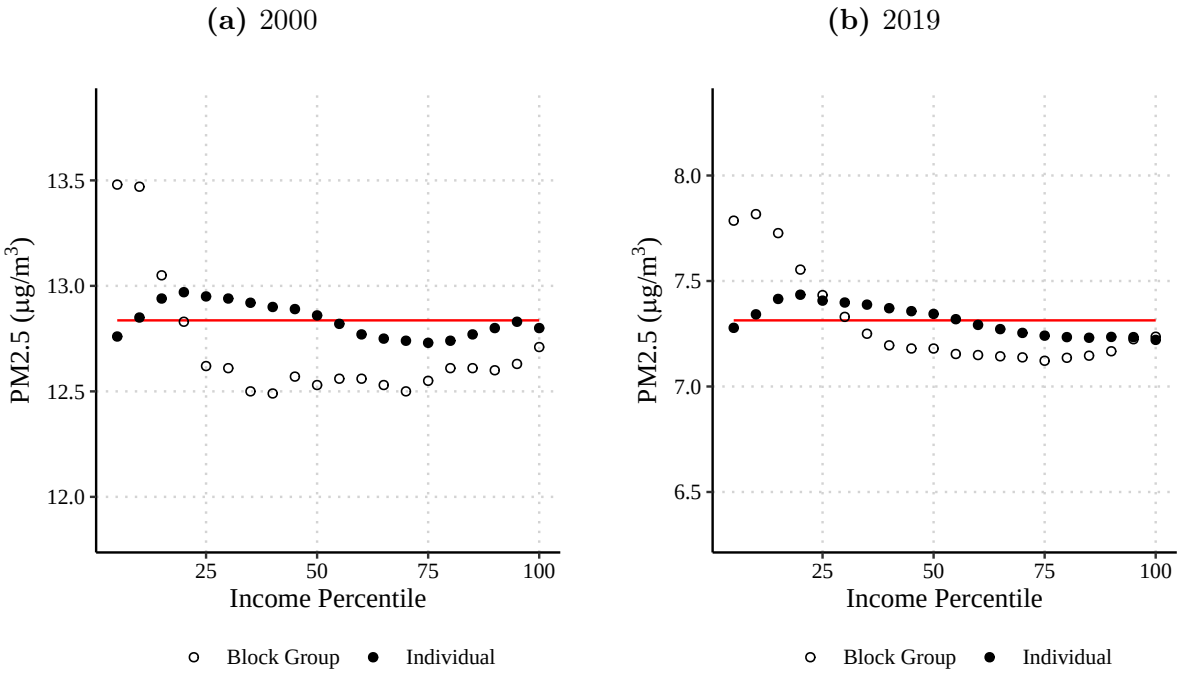
Source: Environmental Impacts Frame Residential History File, 1999-2021, and Cooper et al. (2022).

Notes: See 2 for details on construction.

In addition to documenting trends in pollution exposure by race, we can also explore how pollution exposure varies across other important socio-economic dimensions, such as income. To explore how pollution exposure varies within the income distribution, we attach administrative income information to the individuals in the EIF. We use adjusted gross income from form 1040 as our income concept, and the tax unit as our income sharing unit (assigning the tax unit level income to each member of the tax unit in the EIF). We do this using IRS administrative data from tax years 1999-2020. Because we only observe income for individuals who appear on a 1040, we drop all nonfilers.[17]

[17]Approximately 80-90 percent of the US population appears on a form 1040, although there are some groups less likely to file, including individuals disconnected from the formal economy (who are also less likely to appear in the EIF) and the elderly (a group for which income may not be a useful measure of well-being.)

Figure 10: $PM_{2.5}$ Exposure by Income



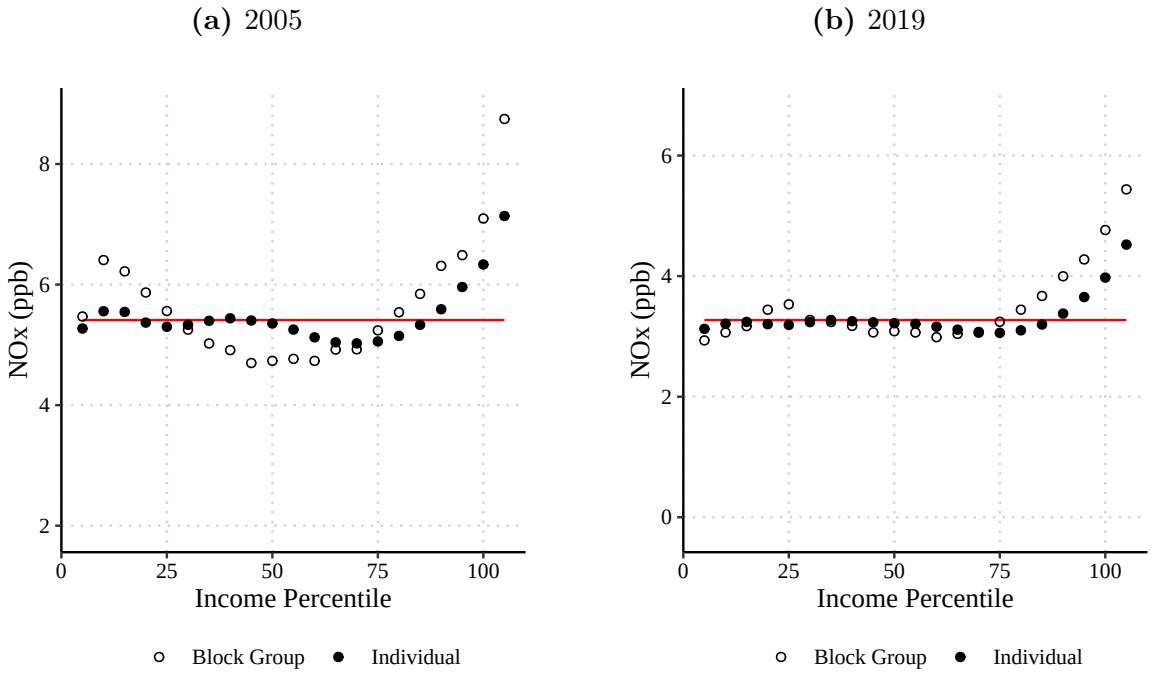
Source: Environmental Impacts Frame Residential History File, 1999-2021, Opportunity Insights Databank, and van Donkelaar et al. (2021). **Notes:** See Section 2 for details on construction.

Figure 10 presents average exposure to $PM_{2.5}$ within income vigintiles for the year 2000 (panel a) and 2019 (panel b). We present two income distributions: the individual tax unit income distribution, and the block group mean income distribution.[18] The purpose of this exercise is to highlight the issues of aggregation bias, discussed further in Colmer et al. (2023c). We see a stark contrast in the two gradients. Using the block group income distribution we see a clear relationship between $PM_{2.5}$ concentrations and average neighborhood income. There is approximately a 1 μ g/m^3 difference in $PM_{2.5}$ between the poorest neighborhoods and the median neighborhood in 2000. A 1 μg/m^3 change in $PM_{2.5}$ is substantial – it is similar in magnitude to the improvement in Black-White $PM_{2.5}$ gaps since 2000, and translates to between \$200–\$1000 in future income losses (Isen et al., 2017; Colmer and Voorheis, 2021). By contrast, we see almost no relationship between $PM_{2.5}$ concentrations

[18]Block Group average income in these figures is calculated from the microdata, so the income concepts are identical and only aggregation differs.

and the individual-level income distribution.

Figure 11: NO_X Exposure by Income



Source: Environmental Impacts Frame Residential History File, 1999-2021, Opportunity Insights Databank, and Cooper et al. (2022). **Notes:** See Section 2 for details on construction.

Income gradients for NO_X exhibit a different patterns and do not appear to suffer from the same aggregation issues as $PM_{2.5}$. Figure 11 shows income gradients for NO_X for 2005 and 2019 and again contrasts individual and aggregate income distributions. Income gradients for NO_X are upward sloping in the top quintile of the income distribution for both the individual and block group income distribution. The individual vs. aggregate relationships do not deviate as much as in the $PM_{2.5}$ case. Again, the differences in these patterns likely emerge as a result of the different different emissions sources, and the fate and transport for the two pollutants.[19] These findings and their causes are examined in much more detail by (Colmer et al., 2023c).

The patterns in people vs. place income gradients across the two pollutants are consis-

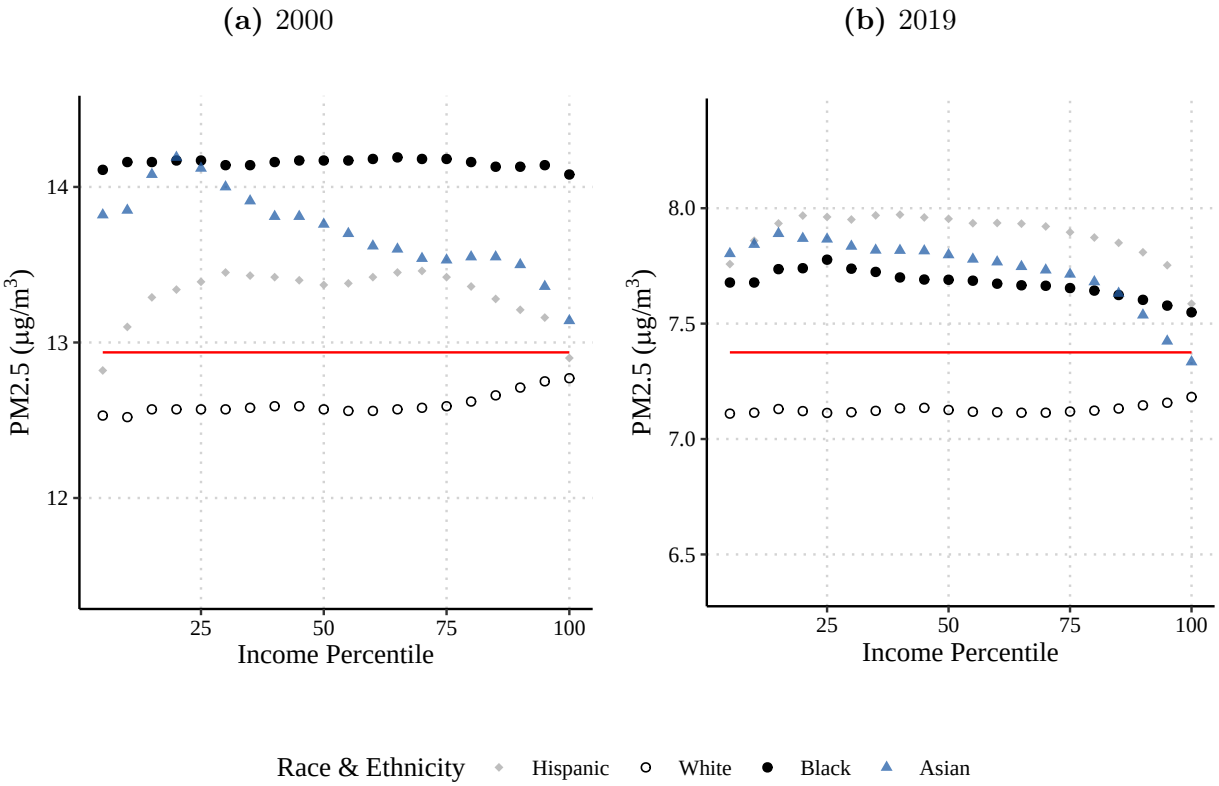
[19] "fate and transport" refer to the processes in which the nature of contaminants change (chemically, physically, or biologically) and where they go as they move through the environment.

tent with differences in how the population distribution and fate and transport of the two pollutants interact. The place gradients for $PM_{2.5}$ are downward sloping in the bottom of the neighborhood income distribution, but flat elsewhere. $PM_{2.5}$ is a more dispersed pollutant, with higher levels in urban areas. It does not necessarily exhibit sharp hotspots. The poorest Census block groups by average income are predominantly in urban areas (and hence have higher $PM_{2.5}$ exposure), but this is not necessarily the case for the poorest people – individuals in the bottom quartile of the income distribution are dispersed across urban, rural and suburban areas. The differences in where poor places and poor people are located are consistent with the different income gradients between people and places for $PM_{2.5}$ that we observe.

The patterns for NO_X income gradients for both people and places are more aligned again because of how fate and transport of NO_X interacts with the population distributions. NO_X is a local pollutant which is relatively unstable, and hence does not have a long lifespan as an ambient pollutant (NO_X is a precursor to ozone and nitrate particulates). Thus NO_X ground level exposure is much less dispersed, and more concentrated immediately near emissions sources – which for NO_X is predominantly from diesel combustion, which in turn is most concentrated in dense areas in city centers. This is consistent with both the individual and block-group income gradients we observe for NO_X – individual and block group income, especially at the top of the distribution are both correlated with density.

With individual-level data it is also possible to explore how exposure varies between multiple dimensions, such as between percentiles of the distribution of income by race (or sex, or education); this contrasts strongly with a place-based approach, which can consider only one margin at a time, e.g., racial composition or the median income of a neighborhood. In the context of air pollution, this is especially important, as neighborhood and individual incomes may diverge substantially, as might individual vs. neighborhood racial demographics.

Figure 12: Income Disparities in $PM_{2.5}$ Exposure by Race and Ethnicity

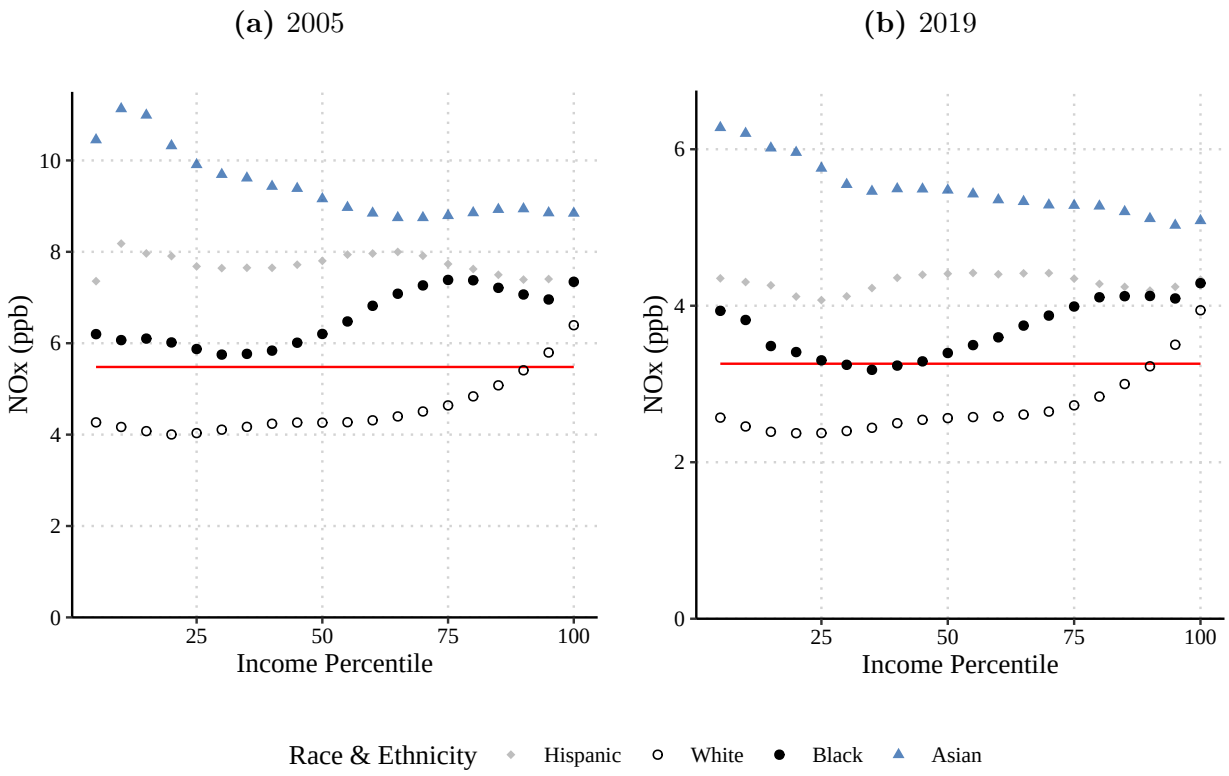


Source: Environmental Impacts Frame Residential History File, 1999-2021, Opportunity Insights Databank, and van Donkelaar et al. (2021). **Notes:** See Section 2 for details on construction. The horizontal red line represents mean $PM_{2.5}$ exposure that year across the four race and ethnicity groups shown in the figure.

Figure 12 presents race-specific income gradients for $PM_{2.5}$, paralleling the structure of earlier figures. Income percentiles in these figures are defined at the national level, so the graphs allow for the comparison of racial gaps between individuals with incomes in the same national income vigintile. We document several important facts: First, non-Hispanic White individuals are exposed to lower levels of pollution in every part of the income distribution; Second, the narrowing of absolute Black-White $PM_{2.5}$ gaps has happened regressively – Black-White gaps narrowed more in the top vigintile of the income distribution than in the bottom vigintile. Third, there is a much more variation in $PM_{2.5}$ for Asian individuals across the income distribution – the Asian-White gap is 1.5 $\mu g/m^3$ in the bottom income

vigintile in 2000, compared to around 0.2 in the top vigintile. Finally, we note that widening Asian-White and Hispanic-White gaps between 2000 and 2019 appear to be driven predominantly by widening gaps in the bottom of the income distribution.

Figure 13: Income Disparities in NO_X Exposure by Race and Ethnicity



Source: Environmental Impacts Frame Residential History File, 1999-2021, Opportunity Insights Databank, and Cooper et al. (2022). **Notes:** See Section 2 for details on construction. The horizontal red line represents mean NO_X exposure that year across the four race and ethnicity groups shown in the figure.

Figure 13 presents race-by-income gradients for NO_X. The upward sloping income gradients seen in Figure 11 appear to be driven by upward sloping gradients within the White income distribution; the income gradient for other groups is flat or downward sloping. In contrast to $PM_{2.5}$, the shrinking racial disparities in NO_X pollution appear to have been distributionally neutral – the narrowing of gaps between 2005-2019 were approximately proportional across the income distribution.

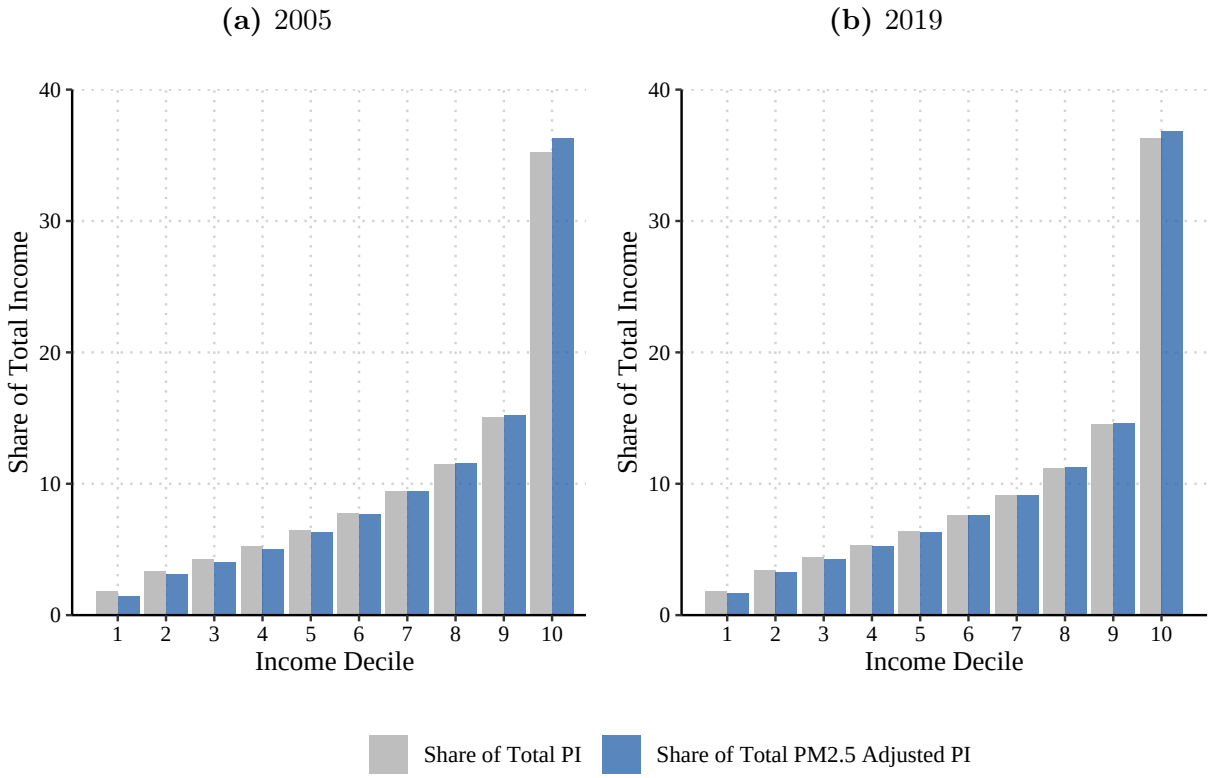
4.2 Distributional Natural Capital Accounting

In this section, we use the individual-level $PM_{2.5}$ income gradients to do some simple national accounting exercises, in the spirit of Muller et al. (2018) and Piketty et al. (2017). Muller et al. (2018) shows that adjusting total GDP by the monetary value of environmental damages produced by this economic activity can substantially change interpretations of longer run trends in economic growth, while Piketty et al. (2017) show that the distribution of economic growth has disproportionately accrued to the top of the income distribution. We thus explore if our individual income gradients can shed light on whether pollution-adjusted income growth has exhibited a different pattern.

To do this, we follow a simplified version of the Muller et al. (2018) approach, applied to the newly released distributional personal income accounts produced by BEA.[20] Specifically, we take a per capita value of the monetary damages of $PM_{2.5}$ exposure from a recent review (The World Bank, 2022) [21] and apply this to the income gradients show in Figure 10 to produce decile-specific aggregate damages due to $PM_{2.5}$ in 2000 and 2019. We subtract these aggregate decile $PM_{2.5}$ damages from the aggregate personal income accruing to each decile, and calculate income shares and inequality statistics of these pollution exposure adjusted personal income amounts. This is visualized in Figure 14.

[20] See https://www.bea.gov/data/special-topics/distribution-of-personal-income for more details.
[21] Around \$150 per person per $\mu g/m^3$ in 2019.

Figure 14: The Distribution of Personal Income, Adjusted for $PM_{2.5}$ Exposure



Source: BEA Distribution of Personal Income Accounts, Environmental Impacts Frame Residential History File, 1999-2021, Opportunity Insights Databank, and van Donkelaar et al. (2021). **Notes:** See Section 2 for details on construction. This Figure shows the distribution of Personal Income by Decile before and after adjusting personal income for the monetary damages of $PM_{2.5}$.

Adjusting for $PM_{2.5}$ exposure exacerbates inequality, consistent with Muller et al. (2018), by increasing the share of income accruing to the top decile and reducing the bottom decile's income share. However, since pollution exposure (and hence damages) has been declining over the period 2000–2019, the trends in pollution adjusted income deviate from the trends in the distribution of personal income. In 2000, the Gini coefficient of Personal Income was 0.432, compared to a Gini coefficient of 0.451 for pollution-adjusted personal income. In 2019, the Gini coefficients were 0.435 and 0.444 for personal income and pollution adjusted income respectively. Although $PM_{2.5}$ exposure exacerbates inequality in the cross-section, improvements in air quality have actually been sufficient to reverse the trend of rising in-

equality in non-adjusted income.

5 Conclusion and Next Steps

This paper describes the creation of a prototype microdata infrastructure – the Environmental Impacts Frame – and showcases one application, documenting how it can be used to expand our understanding of pollution disparties in the United States. While our chosen proof of concept relates to air pollution, the EIF has the potential to fundamentally advance our understanding about any environmental amenities or hazards that can be measured with spatially resolved data. For example, the EIF is an ideal framework for studying the distribution of exposure to, and the consequences of, climate change – increasingly severe wildfires and hurricanes, extreme heat and sea level rise. More broadly, the EIF's longitudinal nature (and incorporation of place of birth information) make it useful for studying dynamic effects, and make it the only population-scale framework with the capacity to seriously answer questions related to environmental migration, sorting, siting, and environmental gentrification.

We have planned several EIF extensions, which we will implement going forward is resources permit and will substantially expand available household information. These include incorporating additional measures of income; worker-firm linkages, which will allow researchers to consider workplace exposures, as well as labor market responses more generally; a historical EIF and family linkages, which will allow researchers to measure environmental exposures earlier in life, determine family structure, and conduct intergenerational analyses; and housing characteristics as well as homeownership, which will improve researchers ability to disentangle exposure from vulnerability. New analyses and extensions will be released as soon as they are available, subject to disclosure avoidance constraints. Our aim is to fully integrate the EIF with all existing Census infrastructure. Importantly, the EIF provides not only a framework for facilitating new research, but also a framework that the Census Bureau can use to enhance existing data products (such as the Community Resilience Es-

timates), and to develop new products and tools describing the relationship between the people, economy, and environment of the United States.

References

Arrow, Kenneth and Anthony C. Fisher, "Environmental Preservation, Uncertainty, and Irreversibility," *The Quarterly Journal of Economics*, 1974, *88* (2).

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

Brock, William A. and Anastasios Xepapadeas, "Valuing Biodiversity from an Economic Perspective: A Unified Economic, Ecological, and Genetic Approach," *American Economic Review*, 2003, *93* (5).

Burke, Marshall, Jonathan Colmer, Cameron Scalera, and John Voorheis, "Wildfire Smoke and $PM_{2.5}$ Disparities in the United States," 2023. Mimeo.

Chakma, Tridevi, Jonathan Colmer, and John Voorheis, "Heat Disparities in the United States," 2023. Mimeo.

Chang, T.Y., J. Graff Zivin, T. Gross, and M. Neidell, "Particulate Pollution and the Productivity of Pear Packers," *American Economic Journal: Economic Policy*, 2016, *8* (3).

_ , W. Huang, and Y. Wang, "Something in the Air: Pollution and the Demand for Health Insurance," *The Review of Economic Studies*, 2018, *85* (3).

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

Colmer, J. and J. Voorheis, "The Grandkids Aren't Alright: The Intergenerational Effects of Prenatal Pollution Exposure," *Mimeo*, 2021.

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

Colmer, Jonathan, John Voorheis, and Brennan Williams, "Air Pollution and Economic Opportunity in the United States," 2023. Mimeo.

_ , _ , **and** _ , "Who Weathers the Storm? The Unequal Effects of Hurricanes in the United States," 2023. Mimeo.

_ , **Suvy Qin, John Voorheis, and Reed Walker**, "The relationship between income, wealth, and pollution exposure: evidence from lottery winners and administrative tax data," 2023. Mimeo.

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

Cooper, M.J., R.V. Martin, and M.S. et al. Hammer, "Global fine-scale changes in ambient NO2 during COVID-19 lockdowns," *Nature*, 2022, *601.*

Currie, J. and M. Neidell, "Air pollution and infant health: what can we learn from California's recent experience?," *Quarterly Journal of Economics*, 2005, *120* (3).

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

Dasgupta, Partha, *The economics of biodiversity: the Dasgupta review.*, Hm Treasury, 2021.

_ **and Geoffrey Heal**, "The Optimal Depletion of Exhaustible Resources," *The Review of Economic Studies*, 1974, *41.*

Davis-Kean, Pamela, Raymond L Chambers, Leslie Davidson, Corinna Kleinert, Qiang Ren, and Sandra Tang, "Longitudinal Studies Strategic Review: 2017 Report to the Economic and Social Research Council," 2017.

Deryugina, T., G. Heutel, H. H. Miller, D. Molitor, and J. Reif, "The mortality and medical costs of air pollution: Evidence from changes in wind direction," *American Economic Review*, 2019, *109* (12).

Fenichel, Eli P. and Joshua K. Abbott, "Natural Capital: From Metaphor to Measurement," *Journal of the Association of Environmental and Resource Economists*, 2014, *1* (1/2).

Ferraro, Paul J., James N. Sanchirico, and Martin D. Smith, "Causal inference in coupled human and natural systems," *Proceedings of the National Academy of Sciences*, 2019, *116* (12).

Finlay, K. and K. Genadek, "Measuring All-Cause Mortality With the Census Numident File," *American journal of public health*, 2021, *111.*

Frank, Eyal and Wolfram Schlenker, "Balancing economic and ecological goals," *Science*, 08 2016, *353.*

Graff Zivin, J. and Neidell, M., "The Impact of Pollution on Worker Productivity," *American Economic Review*, 2012, *102* (7).

Heal, Geoffrey, "Valuing Ecosystem Services," *Ecosystems*, 2000, *3* (1).

IPBES Secretariat, *Scientific outcome of the IPBES-IPCC co-sponsored workshop on biodiversity and climate change* 2021.

IPCC, "Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change," 2021.

_ , "Climate Change 2022: Impacts, Adaptation and Vulnerability. Contribution of Working Group II to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change," 2022.

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

Jbaily, Abdulrahman, Xiaodan Zhou, Jie Liu, Ting-Hwan Lee, Leila Kamareddine, Stéphane Verguet, and Francesca Dominici, "Air pollution exposure disparities across US population and income groups," *Nature*, 2022, *601*.

Liu, Jiawen, Lara P. Clark, Matthew J. Bechle, Anjum Hajat, Sun-Young Kim, Allen L. Robinson, Lianne Sheppard, Adam A. Szpiro, and Julian D. Marshall, "Disparities in Air Pollution Exposure in the United States by Race/Ethnicity and Income, 1990–2010," *Environmental Health Perspectives*, 2021, *129* (12).

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

Muller, N., P. Matthews, and V. Wiltshire-Gordon, "The distribution of income is worse than you think: Including pollution impacts into measures of income inequality," *PLoS ONE*, 2018, *13* (3).

Piketty, Thomas, Emmanuel Saez, and Gabriel Zucman, "Distributional National Accounts: Methods and Estimates for the United States*," *The Quarterly Journal of Economics*, 10 2017, *133* (2).

Schlenker, W. and W.R. Walker, "Airports, Air pollution, and Contemporaneous Health," *The Review of Economic Studies*, 2015, *83* (2).

Secretariat of the Convention on Biological Diversity, *Global Biodiversity Outlook 5 https://www.cbd.int/gbo/gbo5/publication/gbo-5-en.pdf* 2020.

Solow, Robert, "An almost practical step toward sustainability," *Resources Policy*, 1993, *19* (3).

Stiglitz, Joseph, "Growth with Exhaustible Natural Resources: Efficient and Optimal Growth Paths," *The Review of Economic Studies*, 1974, *41.*

The World Bank, "The Global Health Cost of PM2.5 Air Pollution: A Case for Action Beyond 2021," 2022.

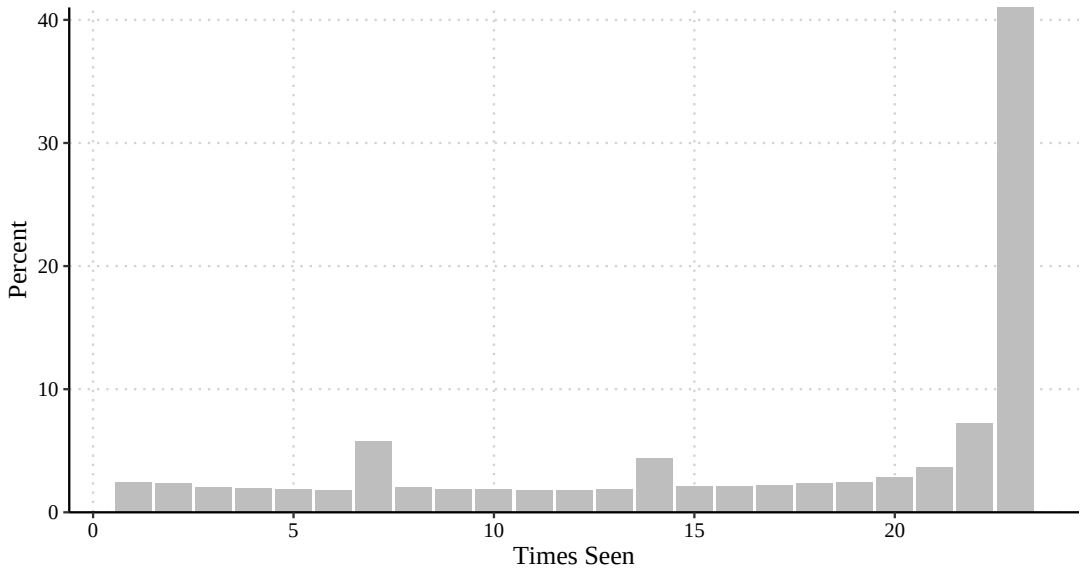
van Donkelaar, Aaron, Melanie S. Hammer, Liam Bindle, Michael Brauer, Jeffery R. Brook, Michael J. Garay, N. Christina Hsu, Olga V. Kalashnikova, Ralph A. Kahn, Colin Lee, Robert C. Levy, Alexei Lyapustin, Andrew M. Sayer, and Randall V. Martin, "Monthly Global Estimates of Fine Particulate Matter and Their Uncertainty," *Environmental Science & Technology*, 2021, *55* (22). PMID: 34724610.

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

Weitzman, Martin L., "On Modeling and Interpreting the Economics of Catastrophic Climate Change," *Review of Economics and Statistics*, 2009, *91* (1).

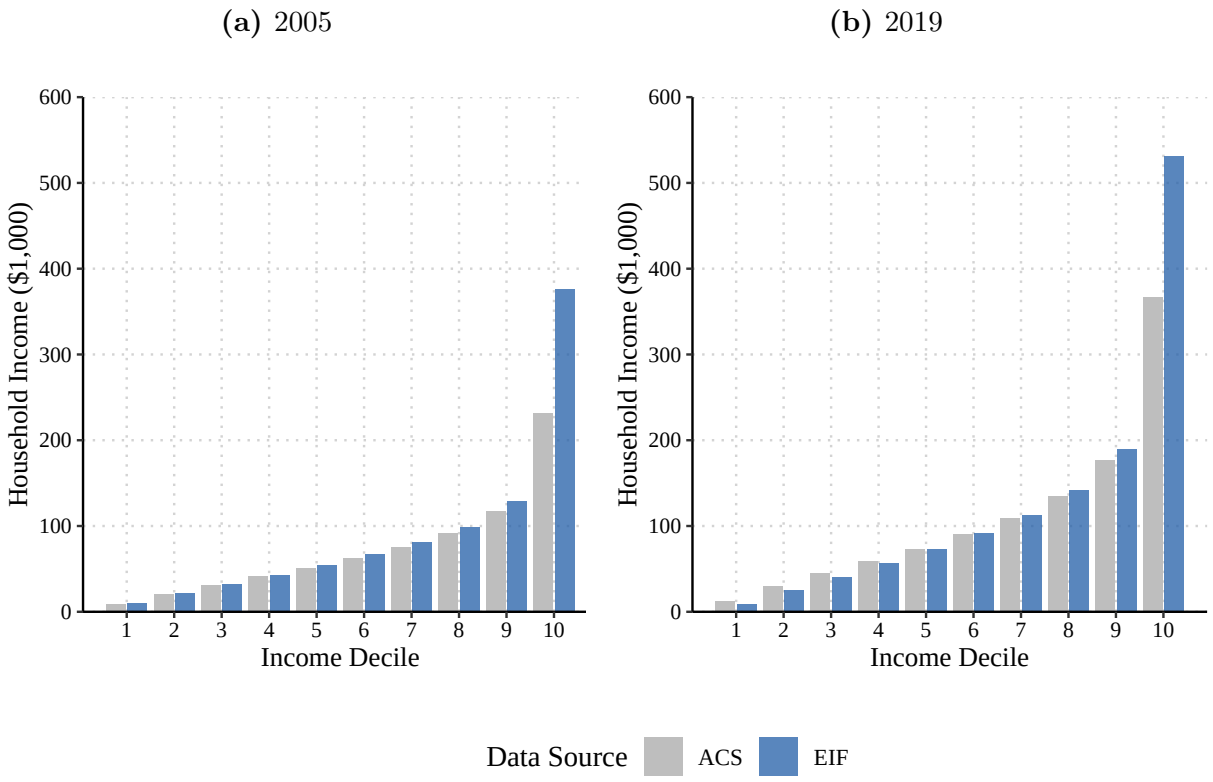
A Appendix Figures

Figure A1: Observation Frequency



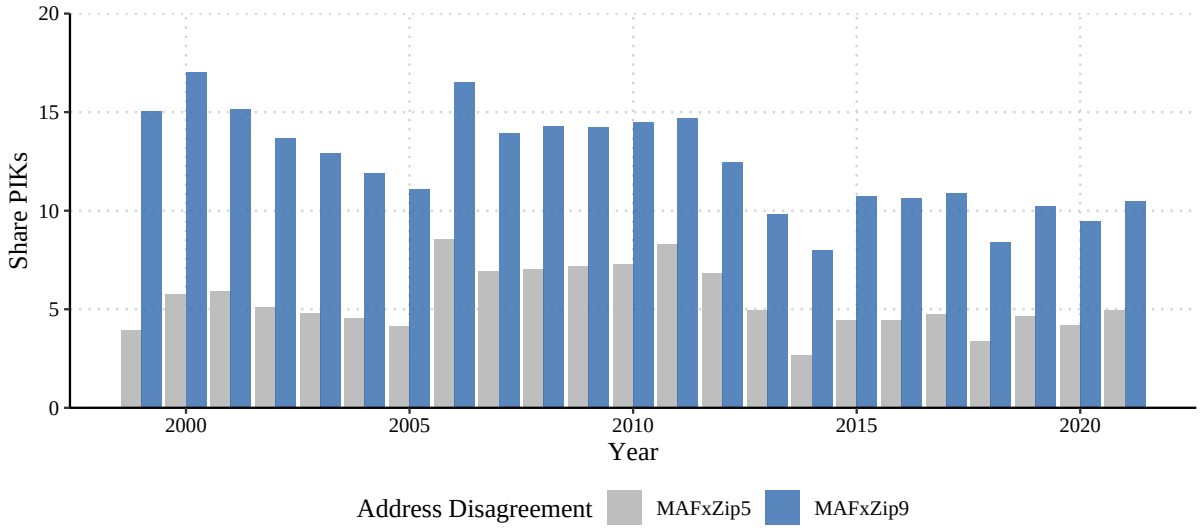
Source: Environmental Impacts Frame Residential History File, 1999-2021. **Notes:** See Section 2 for details on construction.This figure shows the distribution of number of observations per PIK in the EIF.

Figure A2: Income Distribution of Sample



Source: Environmental Impacts Frame Residential History File, 2005-2019, American Community Survey and the Opportunity Insights Databank. **Notes:** See 2 for details on construction. This figure shows the distribution of income in the EIF compared to the ACS.

Figure A3: MAFID and ZIP code Conflicts



Source: Environmental Impacts Frame Residential History File, 1999-2021. **Notes:** See Section 2 for details on construction.This figure shows the the frequency of conflicts between mailing zipcode information and zipcode information from the MAFx.