Environmental disparities in urban Mexico: Evidence from toxic water pollution

Lopamudra Chakraborti[1] and Jay P. Shimshack[2]

December 2021

ABSTRACT: All else equal, disparities in environmental exposure are associated with disparities in health and economic well-being. Here, we combine 9 years of data on toxic water pollution discharges from more than 1,600 industrial facilities across urban Mexico with geographic, economic, and sociodemographic data from ≈ 50,000 Mexican urban block groups. We first show that industrial facilities pollute more in marginalized neighborhoods and in neighborhoods that are becoming more marginalized over time. In contrast, we find no evidence for relationships between toxic water pollution and indigenous race. We then explore channels driving observed exposure disparities. We find evidence that environmental disparities in urban Mexico are associated with collective action, community pressure, and Coasian bargaining. We do not find evidence consistent with political economic or amenity-based sorting mechanisms.

KEYWORDS: environmental justice, environmental disparities, inequality, toxic water pollution

[1] Loyola University Chicago, School of Environmental Sustainability & Quinlan School of Business, 1032 W. Sheridan Road, BVM Hall 411, Chicago, IL 60660, lchakraborti@luc.edu .

[2] University of Virginia, Frank Batten School, 235 McCormick Road, PO Box 400893, Charlottesville, VA 22904-4893. jay.shimshack@virginia.edu .

1. INTRODUCTION

All else equal, disparities in environmental exposure are associated with disparities in health and economic well-being (Cohen et al. 2017, Isen et al. 2017). Environmental disparities have normative welfare implication for fairness, equity, justice, and the interests of future generations (Temkin 2003, Colmer et al. 2020, White House 2021). For these reasons, and others, the unequal distribution of pollution exposure across populations has captured public attention for decades. The media, NGOs, and community groups regularly call attention to correlations between race and poverty and toxic waste, air pollution exposure, deforestation, and land reform (Carruthers 2008, Mohai et al. 2009, Banzhaf et al. 2019). Unequal vulnerability across populations to natural disasters, unsafe drinking water, and the recent global pandemic are further sharpening calls to better understand environmental justice and its underlying mechanisms.

In this paper, we characterize and attempt to explain disparities in toxic water pollution exposure across urban Mexico. We explore associations between urban industrial water pollution and local households' racial composition and wealth status, which we measure using a summary marginalization index. Specifically, we ask, are toxic water pollution discharges higher in more marginalized or higher minority neighborhoods? Are toxic water pollution discharges increasing in neighborhoods that are becoming more marginalized or in neighborhoods with growing minority populations? We then explore the mechanisms that may explain unequal distributions of exposure to toxic water pollution in urban Mexico. Our mechanistic explorations consider channels such as collective action, Coasian bargaining, and community pressure; political economy influences; and Tiebout sorting or "coming to the nuisance" (Banzhaf et al. 2019).

To explore our questions, we combine 9 years of data on toxic water pollution discharges from more than 1,600 industrial facilities across urban Mexico with geographic, economic, and sociodemographic data from ≈ 50,000 Mexican urban block groups. We focus on water discharges of toxic contaminants like arsenic, cadmium, chromium, cyanides, lead, mercury, and nickel because they pose unusually serious human health consequences including cancer, cardiovascular issues, neurological effects, and reproductive abnormalities. We emphasize urban areas of Mexico because most of the world's severe toxic water contamination concerns occur in urbanizing and industrializing areas of middle-income countries like Mexico (Landrigan et al. 2018).

We characterize pollution exposure disparities using raw correlations and nonparametric binned scatterplots that highlight the statistical distribution of pollution across the entire observed range of socioeconomic characteristics (Chetty et al. 2014). We then analyze first-differences and fixed effect models to explore within-location relationships between toxic water pollution, marginalization, and indigenous race, after controlling for trends common to all facilities or all facilities within a region. We highlight sources of variation for our models. We discuss possible limitations for interpreting our results as the effects of marginalization and race on pollution discharges, including threats to causal interpretation from time-varying omitted variables and reverse causality from residential sorting. We then explore mechanisms associated with observed relationships by investigating heterogeneity in pollution / marginalization relationships. We also consider associations between pollution and local migration, residency, and population turnover and what those associations may (or may not) imply for amenity-based sorting.

Our results indicate that industrial facilities' toxic water pollution discharges in urban Mexico are associated with marginalization and increasing marginalization. For example, using within-locality variation over time, we show that a one standard deviation increase in a neighborhood's marginalization index between 2005 and 2013 is associated with a 15 to 40 percent increase in arsenic, cadmium, chromium, cyanides, lead, mercury, and nickel water discharges by local industrial facilities. In contrast, we fail to reject null hypotheses of no statistical relationship between toxic water pollution discharges and indigenous race measures. In terms of underlying channels driving relationships, we find evidence most consistent with collective action and Coasian bargaining. We find no evidence consistent with public political engagement or amenity-based sorting mechanisms.

As public attention to environmental inequality increases, so too does the related literature. We build off earlier studies exploring disparities in pollution exposure and vulnerability to environmental change (Bowen 2002, Ringquist 2005, Mohai et al. 2009, Banzhaf 2012, Agyeman et al. 2016, Narloch and Bangalore 2018, Banzhaf et al. 2019). In other middle-income country settings, Dasgupta et al. (2002), Bayar et al. (2014), Ma et al. (2018), Lome-Hurtado et al. (2019), and Yang and Liu (2019) explore relationships between air pollution exposure and socio-economic factors in Brazil, Turkey, Mexico, and China. Other studies explore local health and economic wellbeing effects of proximity to mining activity in low- and middle-income countries (von der Goltz and Barnwal 2019, Rivera 2020). Additional predecessors examine “pollution havens” and the environmental implications of trade flows between higher and lower income countries (Eskeland and Harrison 2003; Cole 2004; Blackman et al. 2004; Fullerton 2006; Levinson and Taylor 2008; Grineski and Collins 2008; Collins et al. 2009; Grineski et al. 2015).

Relative to this broadly related literature, we make three contributions. First, we provide evidence on pollution disparities related to poverty and race across a large geographic area in a low or middle-income country. Second, we provide evidence on inequities in exposure to toxic water pollution discharges. Third, we investigate the evidence for economic mechanisms driving our observed disparities. For data availability reasons, most empirical studies exploring environmental disparities are set in high-income countries. Studies focused on environmental inequality in low- or middle-income countries typically focus on especially salient air pollution, extractive industry locations, or vulnerability to disasters. Exceptions commonly use data from relatively small geographic areas like a metropolitan area or small transboundary region. Relatively few studies explore heterogeneity in pollution / marginalization relationships in low- or middle-income countries and explore the underlying economic mechanisms.[1]

Our work is most closely related to Chakraborti and Margolis (2017). Chakraborti and Margolis (2017) use similar data from urban Mexico to consider contemporaneous correlations between toxic water pollution and marginalization. We go beyond Chakraborti and Margolis (2017) by exploring relationships between pollution and race, by exploiting graphical evidence and the panel nature of the data to better understand relationships, and by investigating the potential economic channels driving associations. Explaining observed disparities may help inform implications for environmental policy.[2]

2. BACKGROUND & HYPOTHESES

2A. Toxic heavy metals

We study inequities associated with heavy metal contaminants discharged into water. We focus on seven potentially toxic substances: arsenic (As), cadmium (Cd), chromium (Cr), cyanides (CN-), lead (Pb), mercury (Hg), and Nickel (Ni). We investigate heavy metals because large acute or chronic exposures can have serious human health consequences. The Commission for Environmental Cooperation (CEC), the organization largely responsible for establishing the toxic release registry in Mexico, designates our seven pollutants as "special interest" (CEC 2011). The

[1] An exception is Rivera (2020), which documents evidence consistent with residential sorting as rental prices decline near mining activity in Chile, especially for new residents and in areas where pollution is unusually salient.

[2] From a data and methods perspective, Chakraborti and Margolis (2017) consider contemporaneous associations between marginalization and pollution at single snapshots in time. We consider the average relationships between marginalization and pollution over the full period from 2005 to 2013. We also differ by defining toxic water pollution to include both direct and indirect releases.

Lancet Commission on Pollution and Health identifies four of these substances as among the "greatest threats to health" from chemical pollution (Landrigan et al. 2018). Five of the seven substances (As, Cd, Cr, Pb, Ni) are known or reasonably anticipated to be carcinogenetic in humans.[3] Non-cancer health outcomes associated with these substances include: cardiovascular issues ranging from arrhythmia to heart failure; neurological effects including sensory problems and motor signaling disruptions; reproductive issues ranging from infertility to offspring abnormalities; respiratory problems spanning inflammation to respiratory failure; musculoskeletal issues affecting structure and function of bones and muscles; renal problems including kidney disease and kidney failure; and developmental and neurobehavioral problems in newborns, infants, and children. Other health complications associated with one or more of the substances include endocrine (hormonal) disruption, immune system problems, and dermatological issues.

Industrial facilities are the predominant anthropogenic cause of environmental circulation of As, Cd, Cr, CN-, Pb, Hg, and Ni in water. Important local sources include facilities in the energy; chemicals; metals processing; pulp, paper, wood; concrete, stone, clay, glass; food, beverages, tobacco; and electronics sectors. Health consequences of toxic water pollution often remain highly localized around facilities (CEC 2011). As such, the neighborhood-level associations between pollution and socio-demographics explored in this paper are plausible.

2B. RETC, Mexico's toxic pollution registry

Mexico's contaminant release registry is RETC,[4] which is modeled after the US Toxic Release Inventory (TRI) and the Canadian National Pollutant Release Inventory (NPRI) and

[3] See, for example, EPA's Integrated Risk Information System assessments. https://www.epa.gov/iris .

[4] Registro de Emisiones y Transferencias de Contaminantes (RETC), translated as Pollution Release and Transfer Registry). Spurred by provisions of the North American Free Trade Agreement (NAFTA), RETC was authorized under a 2001 amendment to Mexico's General Law of Ecological Equilibrium and Environmental Protection (el Ley General de Equilibrio Ecológico y Protección al Ambiente).

formally implemented in 2004. Like TRI and NPRI, Mexico's RETC is not a regulatory database. Instead, RETC serves to track and inform stakeholders about potentially toxic pollutants processed, transferred, or released into or onto air, water, or land. Reporting is at the facility-level, so toxic pollution can be traced to sources. Reporting is mandatory for specified facilities and chemicals. Like facilities reporting into the TRI and NPRI, RETC facilities may use a variety of methods to measure or estimate their activities and releases including emission factors, mass balance, engineering calculations, stack testing, and direct measurement. Estimation approaches are not recorded in RETC databases.

Despite similarities between RETC, TRI, and NPRI, aspects of toxic pollution tracking differ between the Mexican registry and its North American counterparts. US and Canadian registries track several hundred pollutants each (TRI >600, NPRI >300), while RETC formally tracks 104 pollutants. In practice, considerably fewer pollutants are regularly reported. Additionally, mandatory reporting procedures and thresholds differ across systems. TRI and MPRI define thresholds largely by manufacturing, processing, or otherwise-used (M/P/U) "activities", while RETC thresholds are defined by both M/P/U "activities" and emissions "releases". As a general rule, for toxic water pollutants, RETC reporting thresholds are considerably more stringent (lower) than TRI or MPRI thresholds.[5] Under RETC, facilities must report As, Cd, Cr, Pb, Hg, and Ni discharges if the total M/P/U exceeds 5 kg per year or if the amount discharged to the environment exceeds 1 kg per year. Facilities must report on Cyanide compounds (CN-) if the total M/P/U exceeds 2500 kg per year or the amount discharged to the environment exceeds 100 kg.

Another difference between RETC and its counterparts is that only facilities in 11 industrial

[5] Similarly, facilities in the US and Canada must have 10 employees for required reporting. There is no employee threshold under the Mexican system.

sectors (including energy, chemical, automotive, cement, metals, and petroleum) must report.[6] In Canada and the US, essentially all facilities meeting thresholds are required to report. Due to this distinction, and due to differences in relative scales of economic activity across countries, fewer facilities report into RETC than into the US TRI or Canadian MPRI. During any given year, between 2500 and 4500 facilities will report into RETC. Historically, less than half of facilities reported releases into water. Roughly 30% reported only greenhouse gases and roughly 20% reported on air and land activities / releases without reporting any water activities / releases.

2C. Pathways linking toxic pollution and socio-demographics

Much of our initial empirics focus on first-order environmental disparities hypotheses:

(H1) Pollution from industrial facilities in urban Mexico is associated with more marginalized neighborhoods and in neighborhoods that are becoming more marginalized over time.

(H2) Pollution from industrial facilities in urban Mexico is higher in minority neighborhoods and in neighborhoods with growing minority populations over time.

Potential mechanisms driving the inequities described in H1 and H2 are well explored in the literature.[7] Several theories focus on inequities largely established at the time of facility siting. For example, industrial owner-operators may locate facilities in minority and marginalized neighborhoods due to preferences for discrimination towards poor or minority populations (Becker 1957; Hamilton 1995; Agyeman et al. 2016). Owner-operators may locate facilities near factors of production - like low land prices, access to transportation, or industrial agglomeration – that are correlated with marginalized or minority populations (Wolverton 2009, 2012).

Low income and minority populations may also be disproportionately exposed to industrial pollution at any given point in time, or over time, via other channels including:

---

[6] A technical exception is that all facilities polluting into "national water bodies" must report releases.
[7] See, for example, Mohai et al. 2009, Banzhaf 2012, Banzhaf et al. 2019.

(M1) Community pressure and Coasian bargaining mechanisms. Industrial owner-operators may pollute more when and where community activism directed towards the facility or where necessary Coasian compensation payments to communities may be lower (Coase 1960; Hamilton 1995; Ash and Fetter 2004; Banzhaf 2012; Timmins and Vissing 2017; Banzhaf et al. 2019). A literature suggests that direct community actions and "informal regulation" are especially influential in low and middle-income countries like Mexico, as more formal regulatory pressures are less prominent (Pargal and Wheeler 1996; Dasgupta et al. 2000; Seroa de Motta 2006; Ma 2010; Feres and Reynaud 2012).

(M2) Public political engagement mechanisms. Industrial owner-operators may pursue opportunities to pollute when and where populations are less politically active in order to minimize formal regulatory attention (Hamilton 1993, 1995; Brooks and Sethi 1997; Arora and Cason 1999; Helland and Whitford 2003). Governments can affect facilities' costs of environmental performance via monitoring, enforcement, and legislative pressure (Banzhaf et al. 2019). Evidence suggests that regulator inspections and penalties are associated with political action correlates in the United States (Gray and Shadbegian 2004; Konisky 2009; Shadbegian and Gray 2012).

(M3) Amenity-based sorting mechanisms. In models of residential sorting (Tiebout 1956, Been 1994), wealthier groups move away from polluted areas and poorer individuals move into polluted areas after declines in property values. This "moving to the nuisance" or "environmental gentrification" amenity-based sorting hypothesis has received particular attention in the economics, policy, and law literature for rich countries (Been 1994; Been and Gupta 1997; Banzhaf and Walsh 2008; Gamper-Rabindran and Timmins 2011, and Currie et al. 2015).

Although we make no attempt to contribute novel theory, we follow the literature and note that the mechanisms above generate sensible empirical predictions under reasonable assumptions. Since our study analyzes pollution conditional on the existence and operation of an industrial facility, we do not consider empirical predictions of mechanisms largely operating at the time of establishment siting. Relevant empirical predictions of (M1) – (M3) may include:

(EP1) The strength of relationships between pollution and marginalization (or race) will be associated with shares of renter-occupied housing if community pressure and bargaining (M1) is an important mechanism (Ash and Fetter 1994; Rohe and Stewart 1996). Renter- vs. owner- occupied housing is a standard proxy for the propensity of local populations to engage in community pressure or bargaining.

(EP2) Relationships between pollution and marginalization (or race) will be associated with voter turnout if public political engagement (M2) is an important mechanism (Shapiro 2005). Voter turnout is a standard proxy for public political engagement.

(EP3) Industrial pollution will be positively associated with residential mobility and population turnover in all areas if amenity-based sorting (M3) is an important mechanism. Advantaged households may move out and disadvantaged households may move in where and when pollution is disproportionately high. Turnover may increase as pollution increases.[8]

3. DATA

To explore the causes and consequences of environmental disparities in urban Mexico, we use fine-scale sociodemographic data from the Mexico National Institute of Statistics and Geography's Population and Housing Census and the Mexico National Population Council (Consejo Nacional de Población, CONAPO). We match these sociodemographic data to annual

[8] Although out-migration of advantaged households and in-migration of disadvantaged households may leave total population unchanged, sorting is expected to lower the population share that has lived in the same area for many years.

facility-level toxic pollution discharges data from RETC. Later mechanistic explorations match data to additional sources including voter turnout data from the Federal Election Institute (IFE) and municipal-level migration data and local-level renter share data from the census.

3A. Sociodemographic data

We collected urban AGEB-level data from the 2000, 2005, and 2010 general and "conteo" censuses.[9] AGEBs, for Área Geoestadística Básica Urbana, are the original units of observation. Mexico contains more than 50,000 AGEBs with an average size of ~0.4 square kilometers and a population of approximately 1650 people. For perspective, the average population density in urbanized areas of the US is ~ 800 people per square kilometer. For each urban AGEB, we extract measures like wealth correlates and poverty indicators; educational attainment; health indicators; residency and nationality statistics; indigenous identity; housing tenure; and population density. AGEB-level data facilitate credible local matches between socio-demographics and pollution (Gray et al. 2010).[10]

For ease of interpretation, our main analyses focus attention on relationships between toxic water pollution and two summary sociodemographic statistics. Following Chakraborti and Margolis (2017), our first summary measure is an overall *marginalization index*, which characterizes households' overall wealth status. Direct income measures are unavailable in Mexican censuses and such measures would reflect self-reported annual income rather than more desirable holistic measures of wealth and well-being (CONAPO 2011). Our marginalization index is calculated by the National Population Council based on the first principal components of

[9] General censuses are conducted every ten years for years ending in 0 and "conteo" censuses are conducted every ten years for years ending in 5. Conteo censuses ask fewer questions but still target the entire population.
[10] Although Mexican administrative data also defines rural administrative units, these units differ systematically from urban AGEBs and detailed sociodemographic census data are available only for urban AGEBs. Since sociodemographics are central to our analysis, we study urban AGEBs alone. We reiterate here and elsewhere that urban AGEB-level data imply that we do not study rural areas.

available education, health, household quality, and earnings indicators for the given census (CONAPO 2011, *Data Appendix*). Scores range from small negative numbers to small positive numbers, with more positive index scores indicating greater marginalization and less wealth and well-being. The second summary measure is an *indigenous population* variable. Following official practice of the Mexico National Institute of Statistics and Geography, we identify indigenous race by the fraction of the population over 5 years old that self-reports speaking at least one of 62 native Mexican languages.[11]

3B. RETC data

Using RETC data, we identify facilities emitting toxic substances into Mexican waterways between 2005 and 2013. RETC began in 2004 and data from that year appear incomplete. At the time of our data downloads, RETC data from 2014 onward appeared incomplete. This is likely due to lags associated with populating facility-level information into the national RETC database.

For each RETC facility, we obtain a unique identifier, latitude, longitude, and industrial sector. We use latitude and longitude to establish administrative locations like AGEB (of >50,000), municipality (of >2,000), state (of 32), and region (of 6) (*Data Appendix*). We translate (from Spanish to English) and standardize industrial sector definitions to: energy; chemicals and allied products (including plastics and paints); food, beverages, and tobacco; metals; stone, clay, glass, and concrete (including asbestos); automotive; wood, pulp, and paper; electronics; petroleum and petrochemicals; and other, including services.

For each facility, we obtain self-reported annual pollution discharges, measured in kilograms of emissions, for arsenic (As), cadmium (Cd), chromium (Cr), cyanides (CN-), lead

[11] Indigenous race data are only available from the 2005 Conteo census. The Mexican census also defines race using self-reported cultural identity to identify indigenous populations. This measure is highly correlated (r>.95) with the more "official" linguistic approach to defining race. Our choice has no bearing on any subsequent results.

(Pb), mercury (Hg), and Nickel (Ni). For each contaminant, we analyze total pollution discharges into water defined as the sum of direct emissions and indirect emissions (via sewage systems). We focus on total discharges for two reasons. First, roughly 75 percent of As, Cd, Cr, CN-, Pb, Hg, and Ni plant-by-year water discharge observations represent only direct emissions. Second, the ultimate fate of indirect emissions is not tracked. Observers note that reductions in toxicity from Mexican wastewater treatment plants are highly variable and that emissions reported as indirect in the RETC database result in at least some direct emissions to surface waters (CEC 2011).

3C. Final analysis sample

Where possible, we construct a final dataset for analysis following the insights from the literature on methodological issues and spatial choices in environmental justice scholarship (Baden and Coursey 2002; Baden et al. 2007; Noonan 2008; Gray et al. 2010; Boyce et al. 2016). For each geo-located facility, we use GIS to spatially join the facility to all AGEBs (akin to smaller census block groups) within a fixed radius of the polluting plant. For each facility, we assign a value to each sociodemographic or supplemental measure using the simple unweighted average of the values for that measure over all AGEBs falling within the fixed radius. Our main analysis uses a 1km radius.[12] We choose census-based geographic units because they reflect populations and neighborhood borders more naturally than postal codes or other administrative units (Gray et al. 2010). We choose relatively small radii for the assignment mechanism to reduce potential for ecological inference fallacy (Gray et al. 2010; Banzhaf et al. 2019). We choose the fixed radius approach because formal exposure assessments for RETC facilities have not been conducted and credible mathematical risk assessments for toxic water pollution in Mexico are unavailable. We acknowledge that toxic heavy metal dispersion and transport in water can be complex and that

[12] We later explore robustness to 1.5km and 0.5km radii, since the literature notes the choice of scale can be particularly important in environmental inequality studies (Baden et al. 2007; Noonan 2008).

risk-based approaches to modeling exposure have advantages in many settings (Ash and Fetter 2004; Mohai et al. 2009). In our urban Mexico setting, however, risk-based or other alternatives to fixed radius approaches would need to rely on unusually strong assumptions.

We construct our final facility sample as follows. We first identified all RETC industrial enterprises reporting any data into RETC for our seven toxic water pollutants in the nine years from 2005 and 2013. This exercise yielded 3,250 facility identifiers. We ultimately analyze a subsample of 1,631 distinct facilities that: (a) reported water pollution discharges for at least one of our 7 toxic substances for at least one year during our sample period; (b) had reliable geo-location information in RETC databases; and (c) could be linked to one or more urban block group-level (AGEB-level) units with complete census data. Our final sample does not include facilities in rural areas, facilities lacking latitude and longitude information, and facilities with latitude and longitude coordinates that were implausible.[13] Our last data cleaning step involved consolidating multiple facility identifiers for the same physical plants.[14]

Armed with a final sample of facilities, we validated pollution data. First, we extensively explored missing pollution data (*Data Appendix*).[15] Second, we preprocessed non-zero pollution

---

[13] "Implausible" locations were determined by inputting coordinates into the Mexican Statistical Agency's National Directory and Statistics on Economic Units (DENUE) database. We also include a legitimate facility located offshore.
[14] 87 of the 1631 sample facilities experienced at least one change in ownership or name during the sample period. Changes in facility ownership or name will result in a new RETC identifier for the same physical plant, so our final sample defines facilities by location and not identifier. We lack data on the exact timing and full details of the name or ownership change. To ensure that name or ownership changes are not driving our results, we replicated the analysis dropping the 87 facilities experiencing name or ownership changes from the sample (*Appendix Table 8*). Results are statistically indistinguishable from the main results, although the smaller sample size results in somewhat less precise estimates in some specifications.
[15] Missing data are not uncommon. In our exploration of missing data, we reassuringly do not find a significant relationship between the act of plant reporting and local marginalization or race (*Data Appendix*). Personnel from Mexico's environment ministry do assess recordkeeping and pollution reporting as part of their air, water, and waste programs. During our sample period, there were 273 fines for recordkeeping violations across all RETC industrial facilities with a mean penalty of USD $30.50, a median penalty of USD $9.60, and a maximum penalty of USD $916. We are unable to assess the accuracy of emissions reports themselves. Incentives to misreport actual release magnitudes, at least from a regulatory perspective, are likely low. RETC is not a regulatory database. Reporting unusually high toxic water pollution discharges into RETC cannot and does not result in any direct government penalties, fines, or sanctions. A literature on self-reported pollution emissions suggests that self-reporting can be

data. Like data in other North American toxic registries, some of the RETC pollution values appear inaccurate (CEC 2009). Officials at Mexico's environment ministry (Secretaria de Medio Ambiente y Recursos Naturales, or SEMARNAT) noted occasional errors in their own reviews of RETC data: "[we note] errors in the conversion of units and errors in the selection of the appropriate substance for report (substances with similar names are often interchanged)" (de Eicker et al 2010, p11-12). In the raw data, the maximum value for each of the seven pollutants was 40 to 100 standard deviations larger than the mean.[16] Absent a clear consensus on the treatment of data entry errors and outliers in the related literature, we preprocess the data with 0.5% trimming. After 0.5% trimming, pollution distributions shared a distributional support with independent evaluations of toxic water emissions from RETC, TRI, and NPRI (CEC 2011).[17]

3D. Summary statistics

Figure 1 maps the 1,631 Mexican industrial facilities in our final sample. Sample facilities span urban areas across the entire country. Roughly 39% of facilities are in the center, 31% are in the northeast, 17% are in the west, 5% are in the northwest, 5% are in the south, and 2% are in the southeast.[18] Our sample includes several clusters of facilities in Mexico's larger and more industrial urban areas including Mexico City (Distrito Federal), Guadalajara / Zapopan, Puebla, Reynosa / Matamoros, Juarez, Tijuana, Monterrey, San Luis Potosi, and Aguascalientes.

---

incentive compatible and can be implemented without affecting facilities' reporting incentives, particularly when fines for high reported discharges are low ($0, in this case) (Malik 1993; Kaplow and Shavell 1994; Shimshack 2014).

[16] As an example of extreme outliers, some values in the raw data for both nickel and lead exceeded independent reports of total emissions of those substances from the US, Canada, and Mexico combined. The maximum value for Nickel in the raw data was roughly 100 billion times greater than the median.

[17] Results are also robust to winsorizing or trimming at smaller or larger thresholds. Main result point estimates are similar in magnitude, but less precisely estimated, when using raw pollution data with no preprocessing.

[18] A limitation of our study is that our sample does not include many maquiladoras under state or municipal jurisdiction. To be clear, if a maquiladora meets at least one of the criteria for major toxic polluter it is included in our database. Some maquiladoras are too small to appear in our dataset. Our fixed effect structures (including facility location fixed effects and locality-by-year fixed effects) should reduce bias from possible correlations between unobserved (small) maquiladora locations, our facility locations and pollution discharges, and local socio-demographics..

31% of 1,631 sample facilities are in the chemicals and allied products sector; 16% are in the metals processing sector; 10% are in automotive sector; 7% are in the electronics sector; 7% are in the food, beverage, and tobacco sectors; and 29% are in other sectors. The average facility is located in a neighborhood where 35% of the population lacks access to healthcare; 22% of the population lives in overcrowded housing (defined as more than 2 persons per room); 20% of the population lacks direct access to piped water; 13% of the population has no refrigerator; 6% of the population lives without sanitary drainage; and 1.2% of the population is indigenous.

Mean discharges of Ni, Pb, Cr, Cd, CN-, As, and Hg are 39, 36, 30, 9, 7, 3, and 1 kilograms per year (*Data Appendix*). We note that average toxic water emissions per facility in our dataset are high relative to those reported into the US TRI or the Canadian NPRI. This fact is confirmed in independent evaluations of RETC (CEC 2011) and reinforces that urban Mexico is an unusually high toxic water pollution setting. For all pollutants, discharges are highly variable and right skewed. Coefficients of variation (σ/μ) are roughly 6 or 7. Pollution varies over time as well as space. Temporal "within" standard deviations are typically smaller than cross-sectional "between" standard deviations, but both are large.[19]

In *Figure 2***,** we plot average water pollution discharges for each contaminant across years. We standardize each pollutant (i.e. subtract μ and divide by σ) to visualize contaminants on comparable scales. The key message of *Figure 2* is that per facility releases of As, Cd, Cr, CN-, Pb, Hg, and Ni are roughly 0.1 to 0.2 standard deviations higher in 2013 than in 2005. *Figure 2* also highlights co-movement in toxic water pollutants.[20]

[19] As: σ-within 18.2, σ-between 20.3; Cd: σ-within 40.4, σ-between 65.6; Cr: σ-within 143.3, σ-between 199.4; CN: σ-within 27.0, σ-between 46.6; Pb: σ-within 123.9, σ-between 253.1; Hg: σ-within 5.2, σ-between 5.1; Ni: σ-within 184.9, σ-between 221.5.

[20] Additional summary statistics, including trends in the components of the marginalization index, are presented in the *Data Appendix*. Most components of marginalization are decreasing (improving) over time, although non-monotonically, including: population ages 6-14 not attending school; households without piped water; households

4. ANALYSIS

4a. Empirical approach

We begin our analysis by characterizing correlations between pollution discharges and local marginalization as well as correlations between pollution discharges and race. We care about these basic associations, because – whatever the cause – it is useful to understand if disadvantaged households are exposed to more toxic water pollution discharges than more advantaged households. These raw correlations have normative welfare implication for fairness, equity, and justice. To explore these correlations, we graph binned scatterplots and estimate their associated conditional expectation functions (following Chetty et al. 2014).

For economics and policy considerations, it is useful to understand whether correlations between toxic water pollution discharges and marginalization (or race) may be ultimately driven by omitted variables correlated with both marginalization (or race) and pollution discharges. Consider a spatial omitted variable like the quality of institutions and governance. An area with systematically poor institutions and lax governance might experience more poverty and more pollution. Consider a temporal omitted variable like economic conditions throughout Mexico. Suppose, for example, that urban Mexico experienced economic shocks that simultaneously influenced changes in poverty and toxic water pollution over time. A related source of confounding might involve a temporal omitted variable common to all areas within a specific municipality (or state or region), but not necessarily common to all areas within the full sample. Suppose, for example, that Guadalajara experienced a policy shock (correlated with both poverty and pollution) not experienced by Mexico City, Monterrey, or other municipalities. In these omitted variable cases, documented correlations between marginalization and pollution may be largely driven by

without a refrigerator; mortality rates; and population without access to basic health services. In contrast, households with overcrowding are increasing over time in our sample.

confounding third factors.

To help minimize identification from time-invariant omitted variables and/or omitted variables trending commonly across facilities within a given geospatial area, we estimate panel models using first differences, facility and year fixed effects, or facility and municipality-by-year fixed effects. These approaches identify statistical relationships using only within-group variation. The research design underlying our estimators can be thought of as a comparison of changes in pollution discharges over time for facilities in areas experiencing larger increases in marginalization over time versus changes in pollution discharges over time for facilities in areas experiencing smaller increases in marginalization over time, after controlling for trends common to all facilities across urban Mexico or all facilities within a specific municipality / state / region.

Another important issue for interpreting relationships between pollution and marginalization or pollution and race is reverse causality from residential sorting. If advantaged households move away from pollution nuisances due to environmental disamenities, and/or disadvantaged households move towards pollution nuisances due to lower rents or home prices spurred by environmental disamenities, pollution may drive marginalization rather than vice versa.

From an empirical perspective, three issues related to sorting bear noting. First, in our first-differences or fixed effects models, identification is within-area only. As such, sorting attributable to time-invariant longer-run average differences in pollution across areas does not contribute to the identification. This is not to say that residential sorting based on longer-run average differences in pollution necessarily does not occur; it is certainly possible that advantaged households move away from (or choose not to locate near) a facility because the local area is more polluted on average than another area. However, such sorting does not contribute to statistical identification and our panel estimates should be interpreted as relationships net of sorting based on average differences. Second, sorting attributable to trends in pollution common all facilities within a municipality (and thus common to all facilities within a state or region) does not contribute to identification in our models with municipality-by-year fixed effects. Again, this is not to say that residential sorting based on trends in pollution within a municipality necessarily does not occur; it is certainly possible that advantaged households move away from (or choose not to locate near) a facility because the entire municipality is experiencing increases in pollution over time. However, such sorting does not contribute to statistical identification and our panel estimates should be interpreted as net of sorting based on trends in pollution common to all facilities within a municipality.

Third, the above points notwithstanding, our approaches do not fully eliminate the possibility of reverse causality due to residential sorting. If, for example, advantaged households move away from (or choose not to locate near) a facility because that particular local area is experiencing increasing pollution relative to other localities within its municipality, our results represent the simultaneous effects of marginalization on toxic discharges and toxic discharges on marginalization. In this case, our main results will overstate the effect of marginalization on toxic water pollution discharges.[21]

Finally, we explore empirical evidence for mechanisms. We explore interactions including the effects of voter turnout and renter/owner occupancy on relationships between pollution and marginalization. We explore evidence for relationships between pollution and population turnover, as measured by the share of the local population living in the area for more than 5 years. We note here and elsewhere that mechanistic explorations should be interpreted with some caution, as

[21] As discussed below, we lag marginalization measures as well, so the concern here is more precisely articulated as 'advantaged households move away from a facility because its future pollution may increase disproportionately relative to other locations in the same municipality.'

interacted facility-level characteristics may be correlated with observed and unobserved confounders. In addition, due to data limitations, our proxies fall short of ideal data.

4b. Methods

First, we present binned scatterplots of pollution discharges vs. our marginalization measure. We then present binned scatterplots of pollution discharges vs. our indigenous race measure (% speaking indigenous languages).[22] Binned scatterplots give non-parametric representations of conditional expectation functions across the entire range of socioeconomic measures when the underlying datasets are too large for simple scatterplots to be interpretable (Chetty et al. 2014). Since scatterplots with hundreds or thousands of observations are difficult to interpret visually due to congestion, binned scatterplots group the x variables into equal sized bins and generate scatterplots using each bin's x variable mean and y variable mean for each data point (Chetty et al. 2014). In all scatter plots, we use a default of 20 bins. As a consequence, each bin (and corresponding point on the plot) represents a summary statistic from 5 percent of the data.

Our binned scatterplots include fitted values from a linear prediction of y on x using the true and complete underlying data rather than binned data. We report, for example, coefficients from bivariate regressions of facilities' arsenic discharges on the locations' average marginalization index. We cluster all standard errors at the state-level to allow for arbitrary correlations across facilities in the same state. Presented p-values reflect tests of null hypotheses of no relationships against naive two-sided alternative hypotheses.

Second, we consider panel approaches. First, we run a first differences analysis where (for example) we regress a facility's long-run change in pollution on the long-run change in the marginalization index for that facility's neighborhood. We analyze two aggregate periods: an early

---

[22] Binned scatterplots were developed in Raj Chetty's lab and are implemented using the 'binscatter' STATA command written by Michael Stepner with input from Jessica Laird and Laszlo Sandor.

period (2005-2007) and a late period (2011-2013). We assign census 2000 demographics to observations in the early period and census 2010 demographics to observations in the late period.[23] Thus, we regress the difference between the average 2011-13 pollution and average 2005-07 pollution against the difference between 2010 and 2000 sociodemographic measures. We later consider robustness to alternate lag structures.

We then run a facility-level fixed effects model using all annual pollution data. [24] Our main approach here assigns census 2000 demographics to each facility's annual observations for 2005, 2006, and 2007; census 2005 demographics to each facility's annual observations from 2008, 2009, and 2010; and census 2010 demographics to each facility's annual observations from 2011, 2012, and 2013. We later consider robustness to alternative lag structures.

Noting the equivalence of first differences and group-level fixed effects models when T=2, our baseline panel approaches have similar formal representations. For facility *i* in time period *t*, we regress pollution discharges on the recent lagged marginalization index for the local area, some controls, and plant-level fixed effects:

$$[1] \qquad ln(Pollution)_{it} = \alpha_i + \beta \ Lagged\_Marginalization_{it} + X_{it}\Gamma + \delta_t + \mu_{it} \ .$$

δ_t represents year fixed effects. Controls X include a time-varying, plant-specific local population density measure. In [1], the aggregation and differencing approach defines time t over a combined early period and a combined late period, so $t = 1,2$. The fixed effects model with annual data defines time t over the 9 years spanning 2005-2013, so $t = 1,2, \ldots, 9$.

[23] The aggregation procedure for pollution involves taking the arithmetic mean over all non-missing data for relevant years. The procedure is analogous to common empirical aggregations like converting monthly data to the annual-level.
[24] Advantages of the long-run aggregation and differencing approach include: it captures the fact that sociodemographics can be slow moving; the assignment of pollution to lagged demographics is intuitively appealing, interpretable, and involves few assumptions; and aggregating and averaging data over multi-year periods may alleviate some measurement error issues. Fixed effects panel approaches use all available data but interpretation (due to complex assignment between pollution and socio-demographics, etc.) is more nuanced.

Specifications with municipality-by-year fixed effects, for facility i in municipality m in time-period t, take the general form:

$$[2] \quad ln(Pollution)_{it} = \alpha_i + \beta\ Lagged\_Marginalization_{it} + X_{it}\Gamma + \tau_{mt} + \mu_{it}\ ,$$

where τ_{mt} are municipality-by-year fixed effects.

All panel analyses use common empirical practices. We log pollution since the underlying pollution distributions are restricted to the positive domain and heavily right skewed (*Data Appendix*). We cluster standard errors at the state-level to allow for arbitrary correlations across facilities in the same state. We test a null of $\beta = 0$ against an alternative hypothesis that $\beta > 0$, given theoretical hypotheses and raw correlations predict positive relationships. We run all regressions of the form [1] separately for each of our seven contaminants. Our seven toxic heavy metals are recognized as individually dangerous and we aim to flexibly allow their health impacts, public perception, and awareness to differ.

Finally, we explore evidence on mechanisms. For local renter housing share or municipality-level voter turnout Z_i, regressions with interactions take the form: [25]

$$[3]\ \ln(Pollution)_{it} = \alpha_i + \beta_1\ Lagged\_Marginalization_{it} + \beta_2\ Lagged\_Marginalization_{it} \times Z_i$$
$$+ X_{it}\Gamma + \delta_t + \mu_{it}\ .$$

To explore evidence for sorting, we examine statistical relationships between pollution levels and time varying population turnover at the municipality-level. Ideal data for this purpose would represent individual-level migration flows data. Such data are unavailable in our setting. Given available data, we run univariate cross-sectional regressions of the percent of the population that was a resident of local municipality for at least five years on average pollution emissions in the

[25] Housing share is measured using 1km radii measures constructed from 2000 census data on the fraction of local housing that is renter occupied. Voter turnout data are from the 2006 presidential year at the municipality level from Mexico's Federal Election Institute (IFE). Since these variables are time invariant, any independent effects on pollution are subsumed into the facility fixed effects α-i.

area. More precisely, we run first differences models with long-term resident measures as the dependent variables and lagged local pollution (in logs) as independent variables.[26] We acknowledge that this analysis will not detect all types of Tiebout-style sorting.

6. RESULTS

6a. Relationships between toxic pollution releases and socio-demographics

Figure 3 presents binned scatterplot results for pollution (in kilograms) plotted against the distribution of the local marginalization index. Results show visually that toxic water pollution discharges are strongly associated with the marginalization index for all seven toxic water pollutants. Although we defer interpretation until later analysis, we note that positive associations between pollution and marginalization are typically large in magnitude and statistically significant at conventional levels. We reject the null of no relationship between pollution discharges and marginalization at $\alpha < 10$ percent for cadmium, chromium, lead, arsenic, cyanides, and mercury.

Figure 4 presents the binned scatterplot results for pollution (in kilograms) plotted against the distribution of the indigenous race measure. Graphical results suggest associations between discharges and indigenous race may be possible for some contaminants, but relationships are inconsistent. Associations between pollution and race are not statistically significant at conventional levels and all slope coefficients on the conditional expectation functions are small in magnitude. We fail to reject a null of no relationship between pollution discharges and indigenous race $\alpha < 10$ percent for all studied pollutants. Since *Figure 4* documents no consistent baseline correlations between pollution and race, we omit race outcomes from the analysis that follows.[27]

[26] To do so, our first differences model maintains an internally consistent lag structure by regressing the 2000 vs. 2010 change in share of residents residing in the municipality for at least 5 years on the 2005-07 on the 2008-10 change in pollution (and controls).
[27] We replicated all analyses with indigenous race as an explanatory variable (both with and without marginalization variables) and continued to find null results. Minimum detected effect sizes were typically small; we are powered.

Table 1 presents main regression results for associations between pollution discharges and local marginalization. Panel A summarizes key results from the aggregation and differencing approach to equation [1]. Panel B presents results from our annual data and fixed effects approach to equation [1]. Panel C presents results from specifications with municipality by year fixed effects of the form of equation [2].

In *Table 1,* Panel A, we reject the null of no relationship between changes in pollution and changes in marginalization at $\alpha < 10$ percent for cadmium, chromium, cyanides, lead, mercury, and nickel. We fail to reject a null of no relationship for arsenic at or near conventional levels. In Panel B, we reject the null of no relationship between changes in pollution discharges and changes in marginalization at $\alpha < 10$ percent for cadmium, chromium, cyanides, lead, mercury, and nickel. Again, we fail to reject a null of no relationship for arsenic at or near conventional levels. In Panel C, we reject a null of no relationship at $\alpha < 10$ percent for all seven pollutants.

It is illustrative to consider the variability of main results to empirical approach. For consistency, we treat the results in Panel B, from regressions on the full dataset and including facility and year fixed effects, as our baseline specifications. Coefficients, which we interpret further below, from these specifications are: As: .12, Ca: .26, Cr: .24, CN-:.28, Pb: .18, Hg: .33, Ni: .29. Coefficients from first-differences models are: As: .17, Ca: .20, Cr: .28, CN-:.32, Pb: .32, Hg: .38, Ni: .29. Results are statistically similar, although first-differences results are less precisely estimated, which is not surprising ex-post given they draw on less data. Coefficients from models with municipality-by-year fixed effects are systematically larger: As: .36, Ca: .57, Cr: .53, CN-:.45, Pb: .21, Hg: .29, Ni: .41.

The relationships between (changes in) pollution and (changes in) marginalization documented in *Table 1* are practically large in magnitude. Interpreting the point estimates in Panel

B, while noting log-linear specifications and $\sigma_{marg.}$ = 1.17, reveals that a 1 standard deviation increase in the marginalization index for the neighborhood around a facility results in: a 14.5% increase in arsenic discharges, a 29.8% increase in cadmium discharges, a 27.6% increase in chromium discharges, a 33.2% increase in discharges of cyanides, a 20.9% increase in lead discharges, a 39.1% increase in mercury discharges, and a 33.9% increase in nickel discharges.

Direct comparison of results with other studies is challenging since few studies consider environmental disparities for toxic water pollution. Papers exploring other environmental disparities typically use different measures and metrics of marginalization. Nevertheless, we can directly compare some of our results to those identified in Chakraborti and Margolis (2017). [28] Relative to that paper, we study relationships between pollution and race. We also exploit the panel nature of the data to better understand relationships between pollution and marginalization. Chakraborti and Margolis (2017)'s cross-sectional findings imply that a one standard deviation increase in marginalization around a facility results in a ~ 40%, 20%, 40%, 49%, 24%, 27%, and 23% increase in discharges of As, Cd, Cr, CN-, Pb, Hg, and Ni respectively. Our panel data estimates suggest that a one standard deviation increase in marginalization around a facility results in a ~ 15%, 30%, 28%, 33%, 21%, 39%, and 34% increase in discharges of As, Cd, Cr, CN-, Pb, Hg, and Ni respectively. As such, we find differences in coefficient magnitudes of 13 to 63 percent between the two studies, depending on contaminant. Another innovation relative to Chakraborti and Margolis (2017) is that – later in this section – we consider the economic mechanisms that may drive the positive relationships documented in both studies.

6b. Robustness

[28] Chakraborti and Margolis (2017) report results as the effect of a 1-unit increase in the marginalization index, but these results can be standardized for comparison to the present study. Chakraborti and Margolis (2017) run analyses for three independent snapshots in time but report that preferred results use pollution data from 2011-2013 and marginalization data from 2010.

We explored robustness to defining local neighborhood using 0.5km and 1.5km radii for sociodemographics in place of the 1km radius. Although the 1km radius is common in the literature, we acknowledge that it is somewhat arbitrarily chosen. For all pollutants, cross-sectional associations remain similar in terms of empirical magnitudes and statistical significance across 0.5km, 1km, and 1.5km radii (*Appendix Tables 1 and 2*). Panel data results are more sensitive to empirical choices. For perspective, panel regressions with 0.5, 1, and 1.5km radii yield coefficients: As: .08, .12, -.02; Ca: .25, .26 .12; Cr: .20, .24, .10; CN-:.25, .28, .17; Pb: .18, .18, .09; Hg: .35, .33, .18; Ni: .29, .29, .17. Panel data point estimates are statistically indistinguishable between 0.5km and 1km radii, although the 1km results are estimated more precisely. Panel results from the 1.5km radii analyses are smaller in magnitude and statistically noisier, but relationships between pollution and marginalization generally remain positive (the exception is arsenic). One possible explanation for smaller and less significant estimates at 1.5km radii is that toxic water pollution is highly localized and often difficult to observe. As such, after some threshold, greater radii simply add noise because relationships of interest do not extend far beyond the facility location (i.e. the clearly observable component of pollution in this context).

We estimated the panel data models with alternative lags for socioeconomic data. Our main first difference model assigns 2005-2007 pollution data to census 2000 and 2011-2013 pollution data to census 2010 data. We replicated the first difference model assigning 2005-2007 pollution data to census 2000 socioeconomic data and 2011-2013 pollution data to roughly symmetric census 2005 socioeconomic data (*Appendix Table 3A*). Our main panel models assign pollution data from 2005-2007, 2008-2010, and 2011-2013 to 2000, 2005, and 2010 census data. We replicated the panel analyses (with and without municipality-by-year fixed effects) assigning pollution data from 2005-2009 to census 2000 socioeconomic data and 2010-2013 pollution data to roughly symmetric census 2005 socioeconomic data (*Appendix Table 3B and 3C*). We replicated the panel analysis assigning pollution data from 2005-2009 to census 2000 data and assigning pollution data from 2010-2013 to census 2010 data, rather than using any data from the 2005 conteo census (*Appendix Table 4*). Finally, we replicated the analysis using a full panel with annual sociodemographic data imputed to the year from 2000, 2005, and 2010 census data (*Appendix Table 5*). In all cases, results for Cd, Cr, CN, Pb, Hg, and Ni were statistically indistinguishable from the results in *Table 1*. Coefficient magnitudes were practically similar or smaller. The possible exception is coefficients on arsenic, which appear sensitive to specification here and elsewhere. For perspective, in our main analysis, panel regressions yield coefficients: As: .12, Ca: .26, Cr: .24, CN-:.28, Pb: .18, Hg: .33, Ni: .29. With linear interpolation, analogous panel regressions yield coefficients: As: .06, Ca: .22, Cr: .21, CN-:.23, Pb: .23, Hg: .31, Ni: .24.

We considered alternative fixed effects structures. For perspective, in our main analysis, panel regressions yield coefficients: As: .12, Ca: .26, Cr: .24, CN-:.28, Pb: .18, Hg: .33, Ni: .29. We considered naïve models without facility-level fixed effects (*Appendix Table 6*). With the exception of arsenic, relationships between pollution and marginalization generally remain positive but magnitudes are small. We considered panel models with state-by-year fixed effects instead of municipality-by-year fixed effects (*Appendix Table 7A).* With state-by-year fixed effects, panel regressions yield coefficients: As: .26, Ca: .35, Cr: .39, CN-:.32, Pb: .27, Hg: .40, Ni: .40. For all contaminants, point estimates remain positive and significant but magnitudes are systematically larger than our main results. We considered panel models with industry-by-year fixed effects (*Appendix Table 7B*). Coefficient magnitudes are generally similar to main results. With industry-by-year fixed effects, panel regressions yield coefficients: As: .09, Ca: .23, Cr: .23, CN-:.30, Pb: .15, Hg: .32, Ni: .26.

Taken as a whole, robustness to model specification suggests insights into identification. Results from baseline models with facility-level fixed effects or first differences are economically and statistically similar to one another but systematically larger than naïve models without facility-level fixed effects. As such, time invariant omitted variables correlated with pollution and marginalization at the facility-level may bias relationships between pollution and marginalization downwards towards zero. Industry-by-year fixed effects appear to offer few additional contributions to identification on average, relative to models with facility-level fixed effects. Results from models with state-by-year and especially municipality-by-year fixed effects are economically larger than baseline results on average, suggesting that (1) omitted time-varying trends common to all facilities within a specific geographic area may bias relationships between pollution and marginalization downwards to zero, and (2) baseline results are unlikely to be overestimated due to sorting based on average differences or common trends in pollution across municipalities or states.

6c. Mechanistic exploration results

Figure 5 and *Table 2* (Panels A and B) summarize results of interaction effects relevant to community pressure and Coasian bargaining mechanisms. We find consistent evidence that relationships between pollution and marginalization are stronger in areas with lower shares of housing that are renter occupied. To put the results in *Figure 5* in context, we find that poverty and pollution are unrelated when the fraction of housing that is renter occupied tops the ~80th percentile of the measure. In contrast, a 1 standard deviation increase in marginalization leads to a ~40 to 80 percent increase in toxic pollution discharges when the fraction of housing that is renter occupied falls below its ~5th percentile of the measure.

To the extent that community pressure and bargaining are channels influencing relationships between pollution and marginalization, one expects a significant interaction effect as noted in earlier empirical prediction (EP1). However, at first blush, the sign on the interaction coefficient may seem counterintuitive. In the related literature from high income countries (Ash and Fetter 1994; Rohe and Stewart 1996), areas with low shares of renting and high shares of home ownership are associated with greater community activism. However, in Mexico, "rentals tend to be a solution that becomes more common where land markets are more mature and property rights are better enforced, so that land occupation and informal housing becomes less of an option" (Fay and Wellenstein 2005, pg. 92).[29]

Figure 6 and *Table 2* (Panels C and D) summarize results of interaction effects relevant to public political engagement mechanisms. We find no statistical evidence that relationships between pollution and marginalization are moderated by voter turnout - the prediction of (EP2). Increases in pollution are associated with increases in marginalization, but we regularly fail to reject a null that voter turnout does not influence the strength of the pollution-marginalization relationship for any of our seven toxic water pollutants. Our one statistically significant effect (at the 10 percent level) has the opposite sign of expectations.

Table 3 summarizes results of direct explorations relevant to amenity-based sorting and "moving to the nuisance." For all seven pollutants, we fail to reject a null of no relationship between the average pollution emissions in the area and the percent of the population that was a resident of local municipality for at least five years. Although more marginalized areas generally experience less population turnover, areas experiencing more pollution or larger changes in

[29] In Mexico, homeownership is generally high among the urban poor (Fay and Wellenstein 2005). In our sample, marginalization is negatively correlated with shares of the population renting (*Data Appendix*). Survey data suggest that households that rent in Mexico are younger on average (Fay and Wellenstein 2005). One natural concern, then, is that the detected moderating role that the share of renter occupancy plays in the pollution-marginalization relationships may be picking up age as an omitted factor. We ran regressions with interaction effects defined by the share of the adult population over 60. We fail to detect consistent evidence that this age measure influences relationships between pollution and marginalization.

pollution over time do not experience greater changes in the fraction of population that represents long-term residents of the municipality. The failure to reject is robust across specifications.

Although we fail to find evidence consistent with residential sorting or "moving to the nuisance" empirical predictions (EP3), we acknowledge that our data are imperfect for the task. Our lack of evidence contrasts with a related literature that often does find evidence consistent with sorting. However, the existing evidence on sorting or willingness to pay for environmental improvements comes largely from high income settings (Been 1994; Been and Gupta 1997; Banzhaf and Walsh 2008; Gamper-Rabindran and Timmins 2011, and Currie et al. 2015) and/or settings with visible or well-publicized disamenities like air pollution and extractive industry installations (Rodríguez-Sánchez 2014, Rivera 2020). We have little evidence on public perceptions of ambient water pollution in less developed countries (Chowdhury et al. 2016), save for a few cases of large environmental catastrophes (Aragones et al. 2017). Toxic water pollution is typically odorless and invisible to the naked eye, and Mexican agencies rarely disseminate information on water toxics to the public (Montes 2018). The limited research suggests awareness of ambient toxic water pollution is likely low in less developed countries (Chakraborti et al. 2010; Robles-Morua et al. 2011; Fisher et al. 2017). Although drinking water assessments are somewhat more common, they remain infrequent or nonexistence in all but the largest metropolitan areas of Mexico. It is at least plausible that residential sorting attributable to toxic water pollution is limited.

7. DISCUSSION AND CONCLUSION

In this paper, we document marked relationships between marginalization and toxic water pollution discharges in urban Mexico. A one standard deviation increase in a neighborhood's marginalization score is associated with a roughly 15-40 percent increase in arsenic, cadmium, chromium, cyanides, lead, mercury, and nickel discharges from nearby industrial facilities. In contrast, we fail to reject a null hypothesis of no relationship between indigenous race and industrial toxic water pollution discharges in urban Mexico. Although this latter result contrasts with the modal result from the related rich country literature, race in Mexico differs from U.S. and European contexts.[30] A cautionary note for interpretation is that indigenous populations in Mexico are concentrated and many indigenous subpopulations live outside of the urban areas considered in this study (INEGI 2009).

We investigate economic mechanisms driving the detected relationships between pollution and local marginalization. We find evidence consistent with community pressure and Coasian bargaining mechanisms (M1) driving observed environmental disparities in urban Mexico. In this sense, our mechanistic evidence complements insights from an earlier literature detailing the importance of "informal regulation" in low- and middle- income countries (Pargal and Wheeler 1996). One contribution of this paper is to highlight implications of this "informal regulation" for environmental *disparities*. In contrast to community pressure and bargaining channels, we find limited evidence consistent with public politics (M2) or amenity-based sorting mechanisms (M3).

We note caveats. First, we are only able to observe local socio-demographic data for urban areas. The results of this study do not necessarily apply to rural areas. Second, we observe industrial toxic water pollution discharges. We do not observe ambient water quality. Our results are conditional on the presence of an industrial facility, and we do not compare pollution discharges between areas with and without industrial facilities. Third, we are unable to fully confirm the accuracy of reported discharges. We explore data quality in detail in the *Data Appendix*, but several unknowns persist. Fourth, our empirical design may not fully isolate

[30] Social distinctions concerning race and ethnicity in Mexico are functions of phenotype, ancestry, language, and other factors (Telles 2014). Indigenous populations have varied ethnic and cultural histories, and the nature and strength of racial identity varies decidedly across space and over time.

causality. Although our panel data estimates are net of many types of residential sorting, it remains possible that sorting attributable to localized deviations from municipal trends in pollution could influence identification. In this case, our panel estimates may overstate the true influence of marginalization on industrial facilities' toxic water pollution discharges. Some very specific forms of localized time varying omitted variables could also influence results interpretation. Finally, our mechanistic explorations are descriptive. Our mechanism proxies may be correlated with other factors. Our proxies are also imperfect. For example, municipality-level proxies for investigating sorting fall short of individual-level migration flows data ideally suited for the purpose at hand.

The above caveats notwithstanding, we believe our results, taken as a whole, have implications for scholarship and policy. Regarding research, we still have relatively limited systematic empirical evidence on the causes and consequences of environmental inequality in low- and middle-income countries. Ideally, future studies would carry our line of inquiry further to better record and illuminate mechanisms, population differences in exposure to harms to health, and avoidance behaviors. Regarding policy, our results suggest a direct role for corrective environmental policies to address *disparities* in exposure to toxic water pollution discharges in urban Mexico.[31] In cases common to the literature - where reductions in harm from salient and visible disamenities like air pollution and extractive industry installations may lead to higher rents and home prices, amenity-based sorting, and "environmental gentrification" - corrective policies may disproportionately benefit advantaged populations and ultimately fail to address disparities (Sieg et al. 2014; Grainger 2012; Banzhaf 2012). In cases where sorting may not be a dominant mechanism, like largely invisible and infrequently publicized toxic water pollution in Mexico, corrective policies targeting specific polluting facilities may help reduce disparities.

[31] We acknowledge that formal environmental regulation in low- and middle- income countries faces challenges (Greenstone and Jack 2015).

REFERENCES

Agyeman, J., Schlosberg, D., Craven, L., & Matthews, C. (2016). Trends and directions in environmental justice: from inequity to everyday life, community, and just sustainabilities. *Annual Review of Environment and Resources*, *41*, 321-340.

Aragonés, J. I., Tapia-Fonllem, C., Poggio, L., & Fraijo-Sing, B. (2017). Perception on the Risk of the Sonora River Pollution. *Sustainability*, 9(2), 263.

Arora, S., & Cason, T. N. (1999). Do community characteristics influence environmental outcomes? Evidence from the toxics release inventory. *Southern Econ. Journal*, 691-716.

Ash, M., & Fetter, T. R. (2004). Who lives on the wrong side of the environmental tracks? Evidence from the EPA's risk-screening environmental indicators model. *Social Science Quarterly*, *85*(2), 441-462.

Baden, B. M., & Coursey, D. L. (2002). The locality of waste sites within the city of Chicago: a demographic, social, and economic analysis. *Resource and Energy Econ.* 24(1-2), 53-93.

Baden, B. M., Noonan, D. S., & Turaga, R. M. R. (2007). Scales of justice: Is there a geographic bias in environmental equity analysis? *J Enviro Planning Management*, 50(2), 163-185.

Banzhaf, H.S., ed. (2012). The Political Economy of Environmental Justice. Stanford Univ. Press.

Banzhaf, S., Ma, L., & Timmins, C. (2019). Environmental Justice: The Economics of Race, Place, and Pollution. *Journal of Economic Perspectives*, *33*(1), 185-208.

Banzhaf, H. S., & Walsh, R. P. (2008). Do people vote with their feet? An empirical test of Tiebout. *American Economic Review*, *98*(3), 843-63.

Bayar, A. A., Degirmenci, S., & Kayalica, O. M. (2014). Does Inequality Matter Air Pollution and Health Relationship?: The Turkish Case, *EconoQuantum* Vol. 12. Núm. 1, pp 7-30.

Becker, G. S. (1957). *The economics of discrimination*. University of Chicago press.

Been, V. (1994). Locally undesirable land uses in minority neighborhoods: Disproportionate siting or market dynamics? *Yale Law Journal*, 1383-1422.

Been, V., & Gupta, F. (1997). Coming to the nuisance or going to the barrios? A longitudinal analysis of environmental justice claims. *Ecology law quarterly*, *24*(1), 1-56.

Blackman, A., M. Batz, and D. Evans (2004). Maquiladoras, air pollution, and human health in Ciudad Juárez and El Paso. *RFF Discussion Paper*, RFF Press, Washington DC.

Bowen, W. (2002). An analytical review of environmental justice research: what do we really know? *Environmental management*, 29(1), 3-15.

Boyce, J. K., Zwickl, K., & Ash, M. (2016). Measuring environmental inequality. *Ecological Economics*, 124, 114-123.

Brooks, N., & Sethi, R. (1997). The distribution of pollution: community characteristics and exposure to air toxics. *Journal of Environmental Econ and Management*, *32*(2), 233-250.

Carruthers, D. V. (2008). Environmental justice in Latin America: Problems, promise, and practice. MIT Press.

CEC, Commission for Environmental Cooperation (Montréal, Québec). (2009). *Taking Stock: North American Pollutant Releases and Transfers Vol. 12,* June 2009.

CEC, Commission for Environmental Cooperation (Montréal, Québec). (2011). *Taking Stock: North American Pollutant Releases and Transfers Vol. 13*. March 2011.

Chakraborti, D., et al. (2010). Status of groundwater arsenic contamination in Bangladesh: a 14-year study report. *Water Res.,* 44(19), 5789-802.

Chakraborti, L., & Margolis, M. (2017). Do industries pollute more in poorer neighborhoods? Evidence from toxic releasing plants in Mexico". *Economics Bulletin*, 37(2), 853-870.

Chowdhury, S., Mazumder, M.A.J., Al-Attas, O., & Husain, T. (2016). Heavy metals in drinking water: Occurrences, implications, and future needs in developing countries. *Sci Total Environ.*, 569-570:476-488.

Chetty, R., Friedman, J. N., & Rockoff, J. E. (2014). Measuring the impacts of teachers II: Teacher value-added and student outcomes in adulthood. *American Econ Review*, 104(9), 2633-79.

Coase, R. H. (1960). The Problem of Social Cost. *The Journal of Law and Economics*, *3*, 1-44.

Cohen, A. J., et al. (2017). Estimates and 25-year trends of the global burden of disease attributable to ambient air pollution: an analysis of data from the Global Burden of Diseases Study 2015. The *Lancet*, 389(10082), 1907-1918.

Collins, T. W., Grineski, S. E., & Aguilar, M. D. L. R. (2009). Vulnerability to environmental hazards in the Ciudad Juárez (Mexico)–El Paso (USA) metropolis: a model for spatial risk assessment in transnational context. *Applied Geography*, *29*(3), 448-461.

Currie, J., Davis, L., Greenstone, M., & Walker, R. (2015). Environmental health risks and housing values: evidence from 1,600 toxic plant openings and closings. *American Economic Review*, *105*(2), 678-709.

Cole, M. A. (2004). Trade, the pollution haven hypothesis and the environmental Kuznets curve: examining the linkages. *Ecological economics*, 48(1), 71-81.

CONAPO (2011), Consejo Nacional de Población. Marginalization Index for Federal Entity and Municipalities; Consejo Nacional de Población: Mexico City, Mexico, 2011.

Dasgupta, S., Hettige, H., & Wheeler, D. (2000). What improves environmental compliance? Evidence from Mexican industry. *Journal of Environ Econ Management*, 39(1), 39-66.

Dasgupta, S., R. Lucas, and D. Wheeler (2002). Plant size, industrial air pollution, and local incomes: Evidence from Mexico and Brazil. *Environ. Development Econ.* 7(2), 365-381.

de Eicker, M. O., Ruddy, T. F., Zah, R., & Hurni, H. (2010). Potentials of Latin American pollutant release and transfer registers as a source of local data for environmental assessments. In Annual Meeting of the Int'l Assoc. for Impact Assessment, IAIA Geneva, Switzerland.

Eskeland, G. S., & Harrison, A. E. (2003). Moving to greener pastures? Multinationals and the pollution haven hypothesis. *Journal of Development Economics*, 70(1), 1-23.

Fay, M., & Wellenstein, A. (2005). Keeping a roof over one's head: improving access to safe and decent shelter. The urban poor in Latin America, 61(19), 91.

Féres, J., & Reynaud, A. (2012). Assessing the impact of formal and informal regulations on environmental and economic performance of Brazilian manufacturing firms. *Environmental and Resource Economics*, *52*(1), 65-85.

Fisher, A.T., López-Carrillo, L., Gamboa-Loira, B., & Cebrián, M.E. (2017). Standards for arsenic in drinking water: Implications for policy in Mexico. *J Public Health Pol*, 38(4):395-406.

Fullerton, D. (2006). The economics of pollution havens. Edward Elgar Publishing.

Gamper-Rabindran, S., & Timmins, C. (2011). Hazardous waste cleanup, neighborhood gentrification, and environmental justice: Evidence from restricted access census block data. *American Economic Review*, *101*(3), 620-24.

Grainger, C. A. (2012). The distributional effects of pollution regulations: Do renters fully pay for cleaner air? *Journal of Public Economics*, *96*(9-10), 840-852.

Gray, W., Belova, A., Huang, J., Post, E. & Kemp, D. (2010). Analytical Methods for Assessing the Environmental Justice Implications of Environmental Regulations: Seminar Summary Report. Report to the US EPA under USEPA Contract #EP-W-05-022. Sept 1, 2010.

Gray, W. B., & Shadbegian, R. J. (2004). 'Optimal'pollution abatement—whose benefits matter, and how much? *Journal of Environmental Economics and management*, *47*(3), 510-534.

Greenstone, M., & Jack, B. K. (2015). Envirodevonomics: A research agenda for an emerging field. *Journal of Economic Literature*, *53*(1), 5-42.

Grineski, S. E., & Collins, T. W. (2008). Exploring patterns of environmental injustice in the Global South: Maquiladoras in Ciudad Juárez. *Population Environment*, *29*(6), 247-270.

Grineski, S. E., Collins, T. W., & Aguilar, M. D. L. R. (2015). Environmental injustice along the US–Mexico border: residential proximity to industrial parks in Tijuana, Mexico. *Environmental Research Letters*, *10*(9), 095012.

Hamilton, J. T. (1993). Politics and social costs: estimating the impact of collective action on hazardous waste facilities. *The RAND journal of economics*, 101-125.

Hamilton, J. T. (1995). Testing for environmental racism: Prejudice, profits, political power? *Journal of Policy Analysis and Management*, *14*(1), 107-132.

Helland, E., & Whitford, A. B. (2003). Pollution incidence and political jurisdiction: evidence from the TRI. *Journal of Environmental Economics and management*, *46*(3), 403-424.

INEGI (2009) "Perfil socioeconómico de la población que habla la lengua indígena", National Institute of Statistics and Geography, Aguascalientes, Mexico.

Isen, A., Rossin-Slater, M., & Walker, W. R. (2017). Every breath you take—every dollar you'll make: The long-term consequences of the clean air act of 1970. *Journal of Political Economy*, 125(3), 848-902.

Kaplow, L., & Shavell, S. (1994). Optimal law enforcement with self-reporting of behavior. *Journal of Political Economy*, 102(3), 583-606.

Konisky, D. M. (2009). Inequities in enforcement? Environmental justice and government performance. *Journal of Policy Analysis and Management*, *28*(1), 102-121.

Landrigan, P. J., Fuller, R., Acosta, N. J., Adeyi, O., Arnold, R., Baldé, A. B., ... & Chiles, T. (2018). Lancet Commission on pollution and health. *The Lancet*, *391*(10119), 462-512.

Levinson, A., & Taylor, M. S. (2008). Unmasking the pollution haven effect. *International economic review*, 49(1), 223-254.

Lome-Hurtado, A., Touza-Montero, J. & White, P.C.L. 2020. Environmental Injustice in Mexico City: A Spatial Quantile Approach. *Exposure and Health*, 12: 265–279

Ma, C. (2010). Who bears the environmental burden in China—An analysis of the distribution of industrial pollution sources? *Ecological Economics*, *69*(9), 1869-1876.

Ma, J., Liu, B., Mitchell, G., & Dong, G. (2019). A spatial analysis of air pollution and environmental inequality in Beijing, 2000–2010. *Journal of Environmental Planning and Management*, 62(14), 2437-2458.

Malik, A. S. (1993). Self-reporting and the design of policies for regulating stochastic pollution. *Journal of Environmental Economics and Management*, 24(3), 241-257.

Mohai, P., Pellow, D., & Roberts, J. T. (2009). Environmental justice. *Annual Review of Environment and Resources*, *34*, 405-430.

Montes, S. (2018). A River of Death. Sierra & Tierra, September.

Narloch, U. & Bangalore, M. (2018). The multifaceted relationship between environmental risks and poverty: new insights from Vietnam, *Environment and Development Economics*, Volume 23, Special Issue 3: Poverty and Climate Change, pp. 298 - 327

Noonan, D. S. (2008). Evidence of environmental justice: A critical perspective on the practice of EJ research and lessons for policy design. *Social science quarterly*, 89(5), 1153-1174.

Pargal, S., & Wheeler, D. (1996). Informal regulation of industrial pollution in developing countries: evidence from Indonesia. *Journal of political economy*, *104*(6), 1314-1327.

Ringquist, E. J. (2005). Assessing evidence of environmental inequities: A meta-analysis. Journal of Policy Analysis and Management: *J Public Policy Analysis Manage*, 24(2), 223-247.

Rivera, N.M. 2020. Is Mining an Environmental Disamenity? Evidence from Resource Extraction Site Openings. *Environmental and Resource Economics* 75 (3), 485-528.

Robles- Morua, A., Halvorsen, K.E., & Mayer, A.S. (2011) Waterborne Disease-Related Risk Perceptions in the Sonora River Basin, Mexico, *Risk Analysis*, 31, 866–876.

Rodríguez-Sánchez, J.I. (2014). Do Mexicans care about air pollution? *Lat Am Econ Rev* 23, 9.

Rohe, W. M., & Stewart, L. S. (1996). Homeownership and neighborhood stability. *Housing Policy Debate*, *7*(1), 37-81.

Seroa da Motta, R. (2006). Analyzing the environmental performance of the Brazilian industrial sector. *Ecological Economics*, *57*(2), 269-281.

Shadbegian, R. J., & Gray, W. B. (2012). Spatial Patterns in Regulatory Enforcement. *The Political Economy of Environmental Justice*, 225.

Shapiro, M. D. (2005). Equity and information: Information regulation, environmental justice, and risks from toxic chemicals. *Journal of Policy Analysis and Management*, 24(2), 373-398.

Shimshack, J. P. (2014). The economics of environmental monitoring and enforcement. *Annu. Rev. Resour. Econ.*, 6(1), 339-360.

Sieg, H., Smith, V. K., Banzhaf, H. S., & Walsh, R. (2004). Estimating the general equilibrium benefits of large changes in spatially delineated public goods. *Intl Econ Rev*, *45*, 1047-77.

Telles, E. (2014). *Pigmentocracies: Ethnicity, race, and color in Latin America*. UNC Press.

Temkin, L. S. (2003). Equality, priority or what? *Economics & Philosophy*, 19(1), 61-87.

Tiebout, C. M. (1956). A pure theory of local expenditures. *J Political Econ.*, *64*(5), 416-424.

Timmins, C., & Vissing, A. (2017). *Environmental Justice and Coasian Bargaining: The role of race and income in lease negotiations for shale gas*. Duke University Working Paper.

von der Goltz, J. & Barnwal, P. (2019). Mines: The local wealth and health effects of mineral mining in developing countries, *Journal of Development Economics* 139 (2019) 1–16.

White House. 2021. "Modernizing Regulatory Review"; https://www.whitehouse.gov/briefing-room/presidential-actions/2021/01/20/modernizing-regulatory-review/

Wolverton, A. (2009). Effects of socio-economic and input-related factors on polluting plants' location decisions. *The BE Journal of Economic Analysis & Policy*, *9*(1).

Wolverton, A. (2012). The role of demographic and cost-related factors in determining where plants locate. *The Political Economy of Environmental Justice*.

Yang, T. & Liu, W. (2019). Does air pollution affect public health and health inequality? Empirical evidence from China, *Journal of Cleaner Production* 203 (2018) 43-52.

Fig. 1. Map of sample facilities. Our 1,631 sample industrial facilities span urban areas across all of Mexico. Clusters occur in or near Mexico City (Distrito Federal), Guadalajara / Zapopan, Puebla, Reynosa / Matamoros, Juarez, Tijuana, Monterrey, San Luis Potosi, and Aguascalientes.

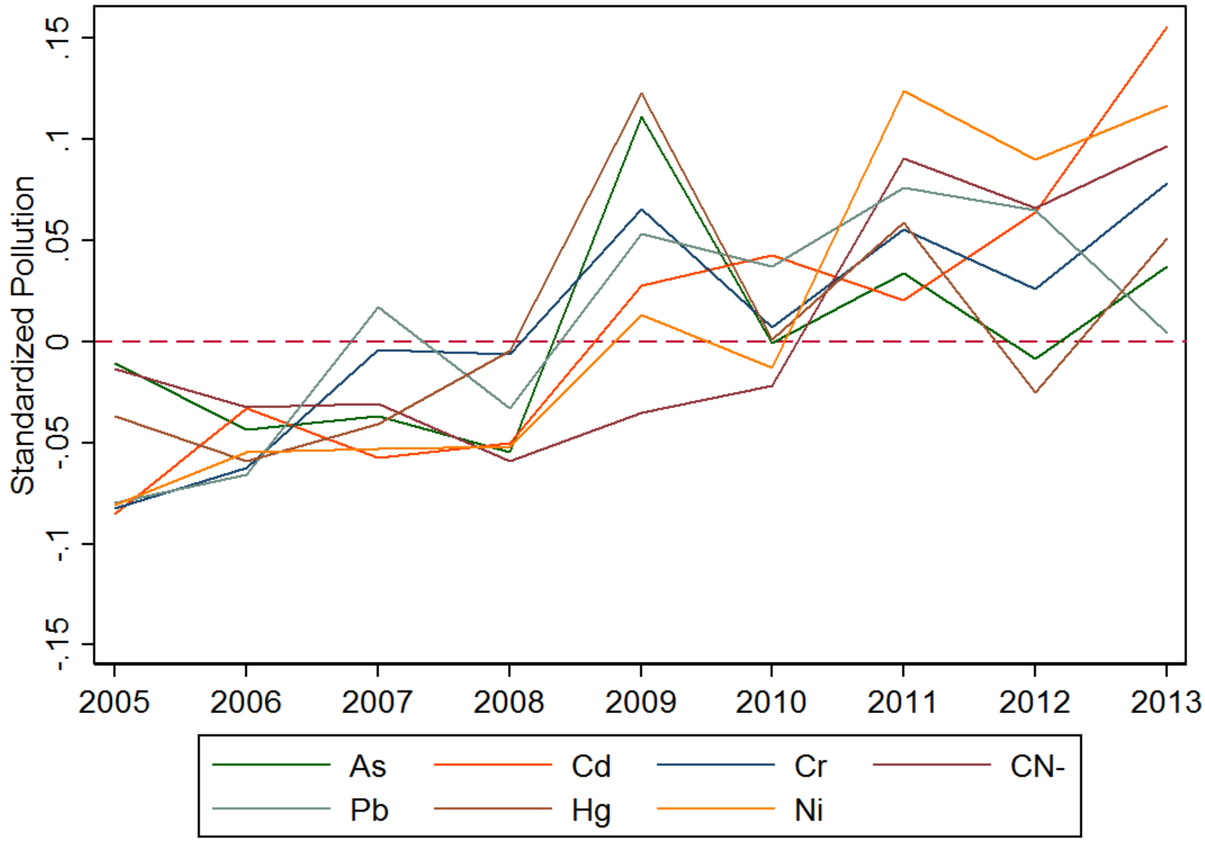
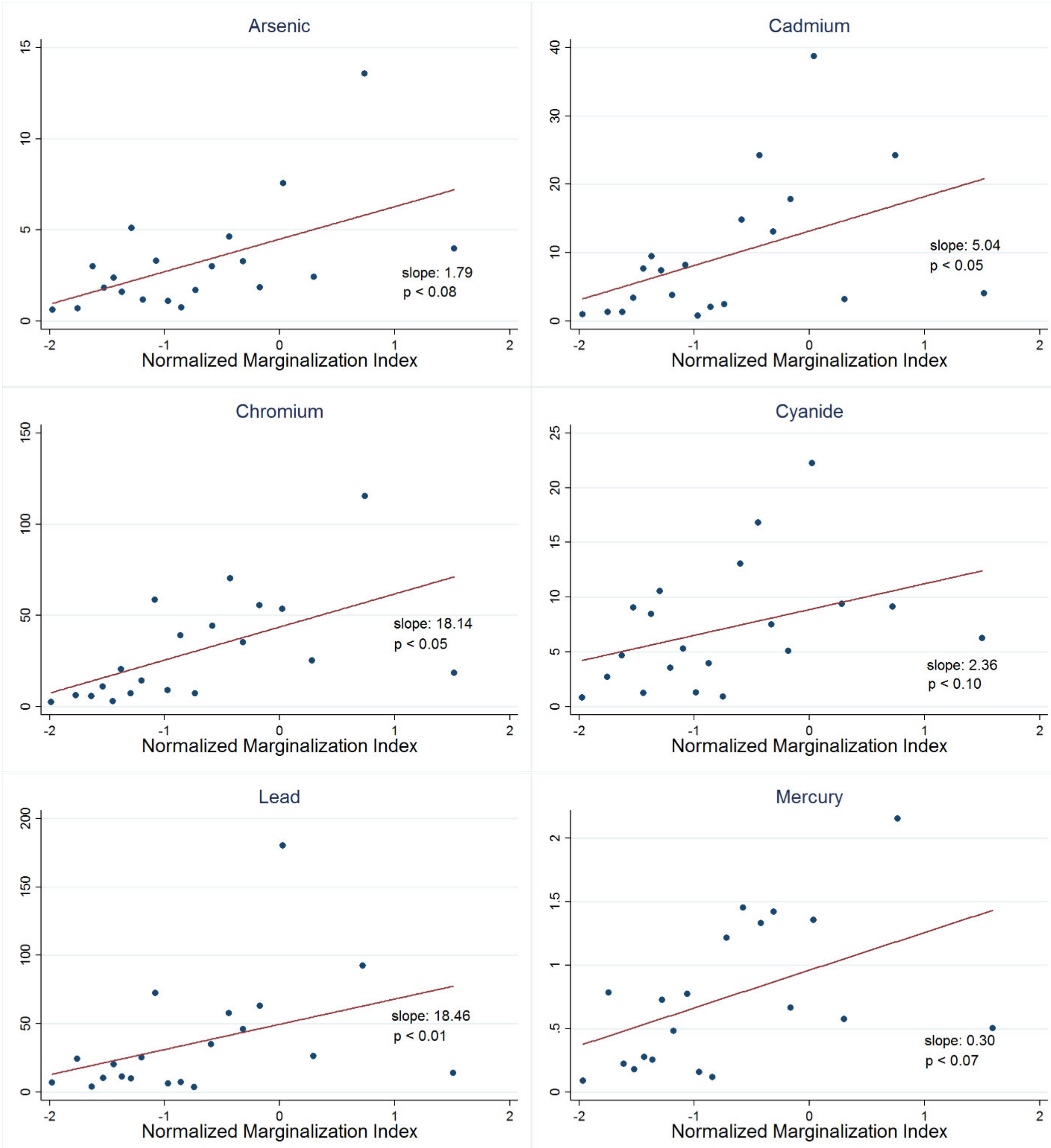


Fig. 2. Trends in toxic releases into water, by pollutant. We standardize pollution discharges in the usual way (subtract out the mean and divide by the standard deviation) in order to present all seven pollutants in the same figure. Pollutants co-move. Reported pollution is increasing over time (although not monotonically).



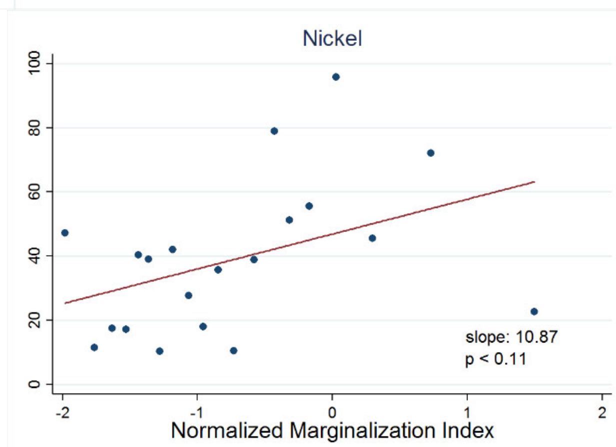
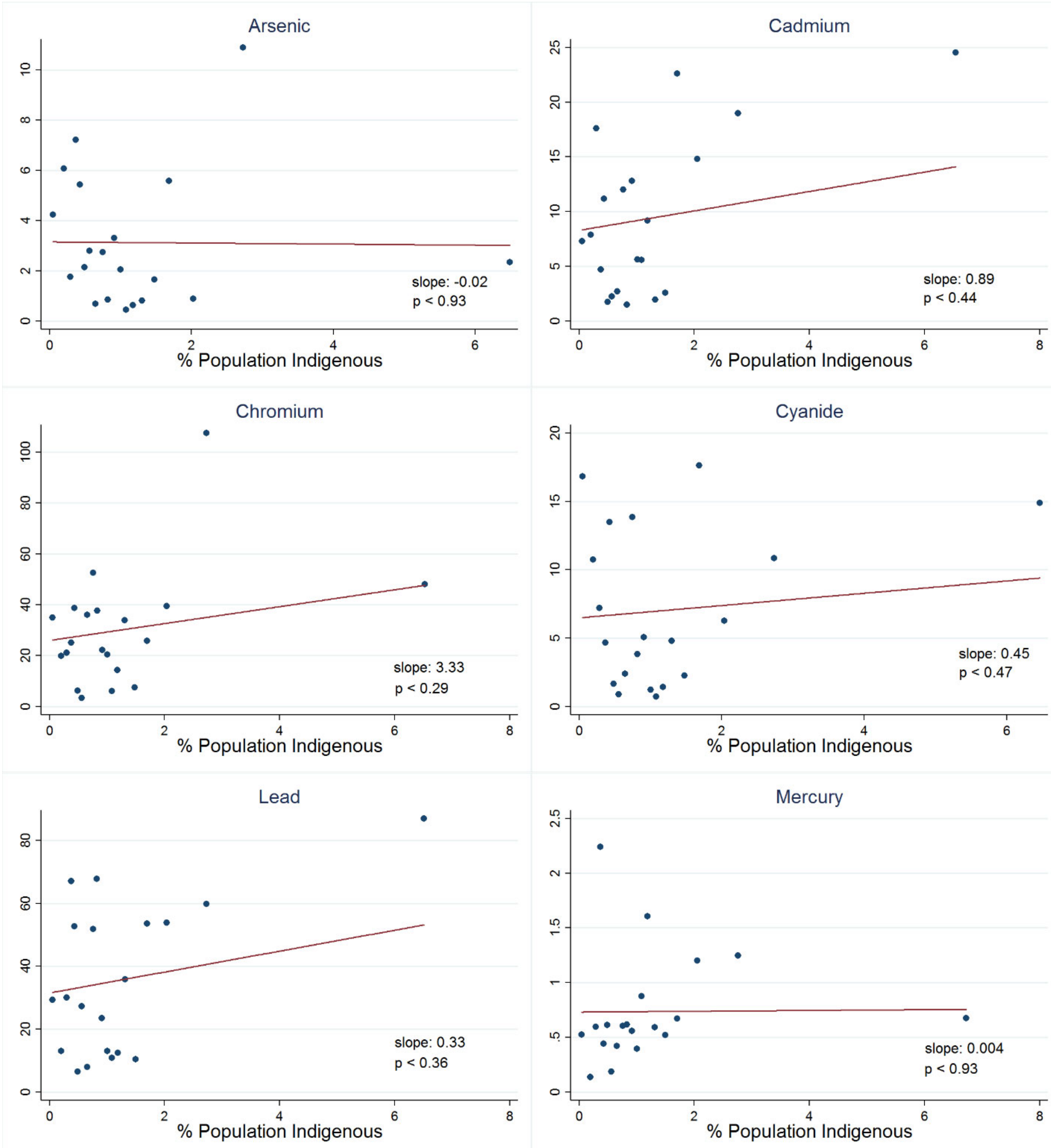


Fig. 3. Binned scatter plots: pollution vs. marginalization. We document positive associations between facilities' local area marginalization scores and toxic pollution. Binned scatterplots summarize relationships and non-parametrically represent conditional expectation functions when the underlying datasets are too large for visualizing full scatterplots. Fitted value regression lines are based on the true underlying data, not on the binned data. Slope coefficients and p-values are overlaid on the figure for each pollutant.



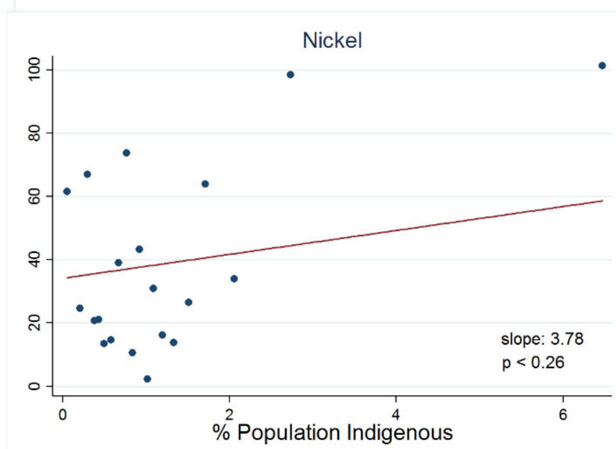
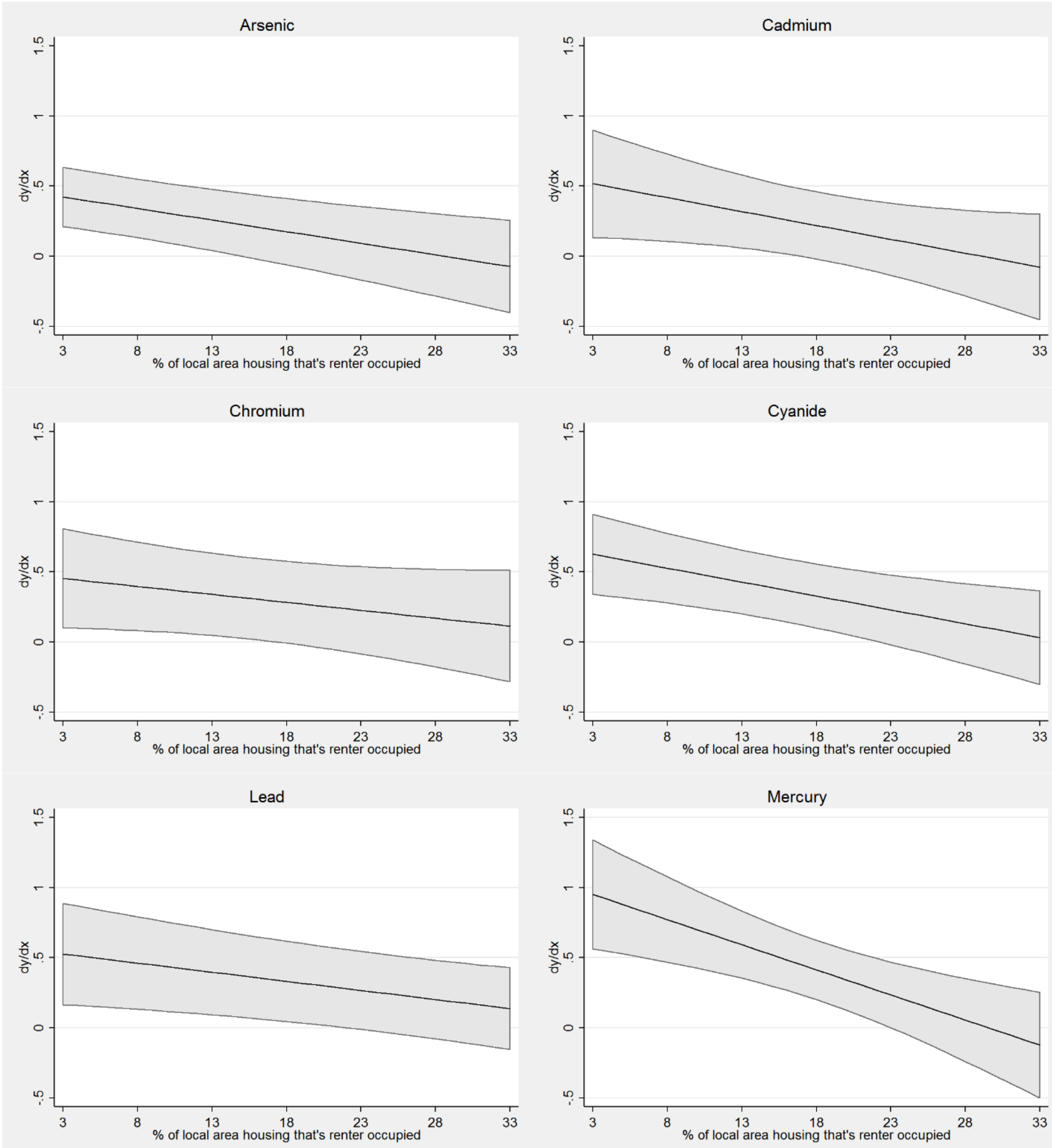


Fig. 4. Binned scatter plots: pollution vs. indigenous race. We fail to reject a null of no association between facilities' local area racial composition and toxic pollution. Binned scatterplots summarize relationships and non-parametrically represent conditional expectation functions when the underlying datasets are too large for visualizing full scatterplots. Fitted value regression lines are based on the true underlying data, not on the binned data. Slope coefficients and p-values are overlaid on the figure for each pollutant.



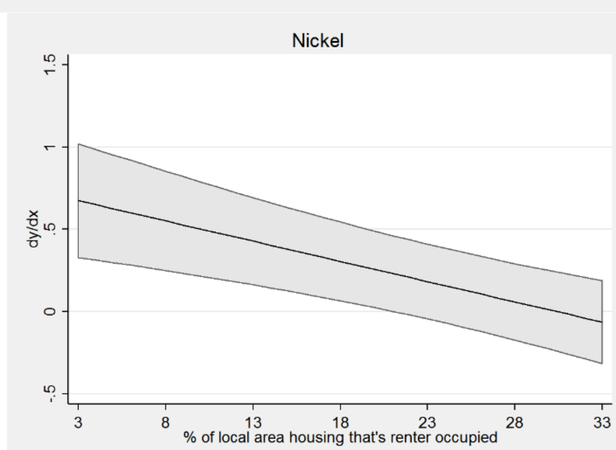
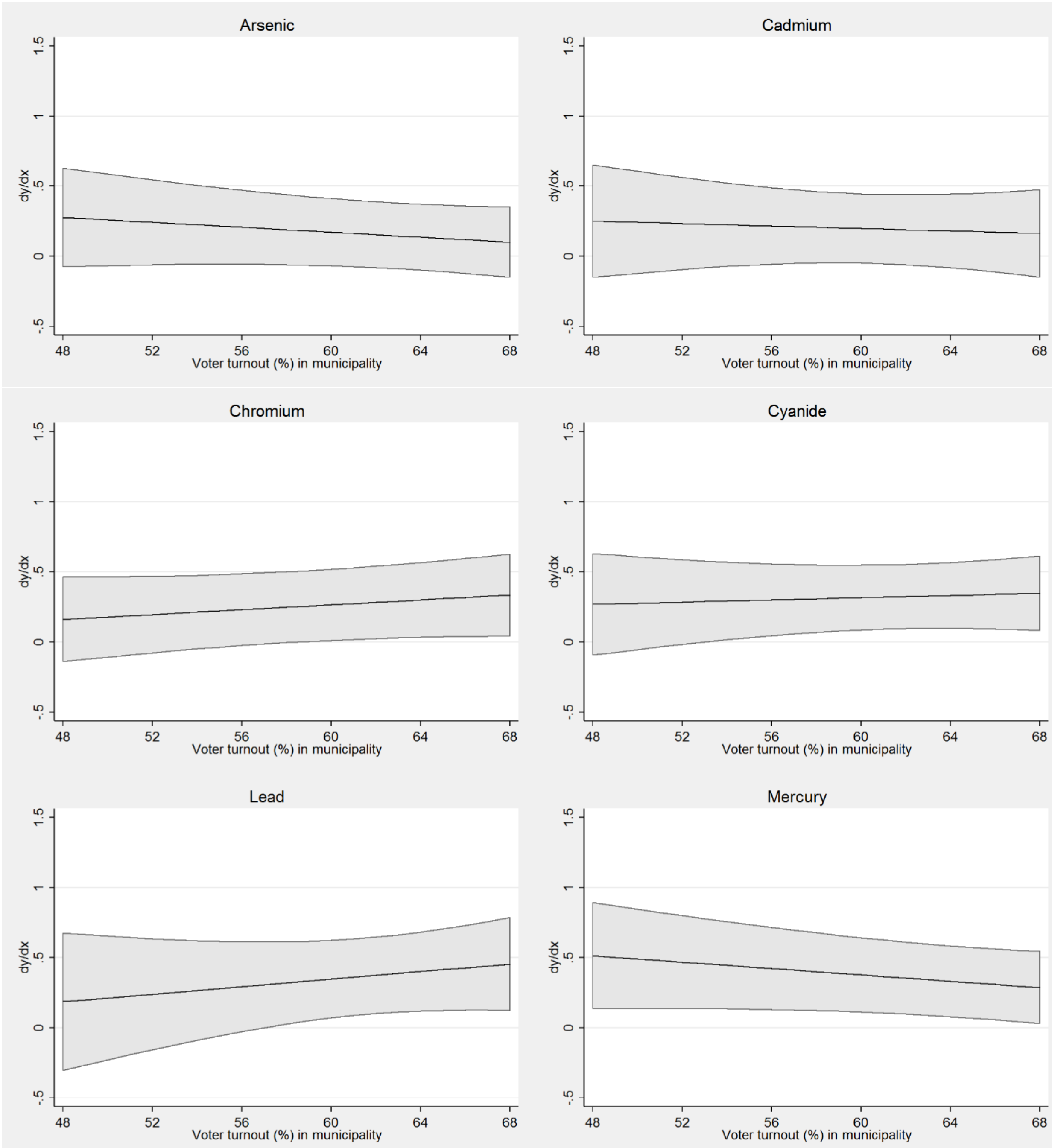


Fig. 5. Renter-occupied housing as moderator of pollution - marginalization relationships. The graphs display the effect of marginalization on pollution at various points in the distribution of renter-occupied housing shares. Although pollution increases with marginalization (dy/dx > 0), the relationship between pollution and poverty is stronger in areas with lower shares of housing that are renter occupied (statistically so for the majority of pollutants).



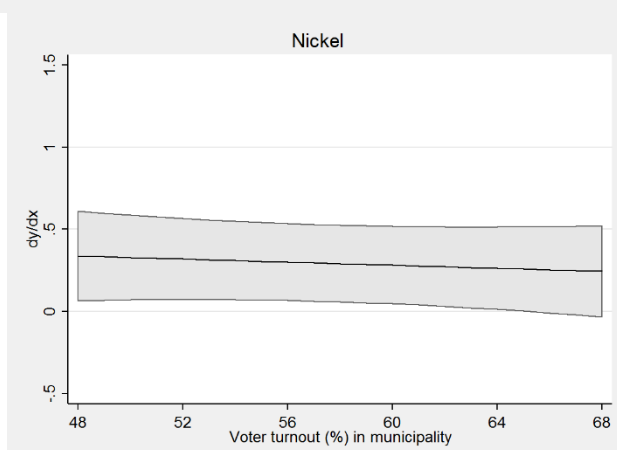


Fig. 6. Voter turnout as a moderator of pollution - marginalization relationships. The graphs display the effect of marginalization on pollution at various points in the distribution of voter turnout. Although pollution increases with increased marginalization (dy/dx > 0), voter turnout does not influence the strength of the relationships.

Table 1. Main Results: Pollution vs. marginalization

Panel A: Panel regressions: First differences, facility by period data							
DEP.VAR: The log of:	As	Cd	Cr	CN-	Pb	Hg	Ni
Marginalization Index	0.17	0.20*	0.28*	0.32**	0.32**	0.38**	0.29**
	(0.14)	(0.15)	(0.17)	(0.14)	(0.17)	(0.15)	(0.14)
Population density	X	X	X	X	X	X	X
Facility FE	X	X	X	X	X	X	X
# observations	779	758	757	783	845	716	869

Panel B: Panel regressions: Full panel, facility by year data							
DEP.VAR: The log of:	As	Cd	Cr	CN-	Pb	Hg	Ni
Marginalization Index	0.12	0.26**	0.24*	0.28**	0.18*	0.33**	0.29**
	(0.11)	(0.14)	(0.15)	(0.10)	(0.12)	(0.12)	(0.12)
Population density	X	X	X	X	X	X	X
Year FE	X	X	X	X	X	X	X
Facility FE	X	X	X	X	X	X	X
# observations	4,250	4,116	4,148	4,277	4,470	4,033	4,483

Panel C: Full panel regressions, municipality with year interactions							
DEP.VAR: The log of:	As	Cd	Cr	CN-	Pb	Hg	Ni
Marginalization Index	0.36**	0.37**	0.53**	0.45**	0.21*	0.49**	0.41**
	(0.16)	(0.22)	(0.27)	(0.15)	(0.16)	(0.18)	(0.19)
Population density	X	X	X	X	X	X	X
Municipality X Year FE	X	X	X	X	X	X	X
Facility FE	X	X	X	X	X	X	X
# observations	4,250	4,116	4,148	4,277	4,470	4,033	4,483

NOTES: Each column reports slope coefficients from a panel regression with toxic water pollution discharges (in logs) as the dependent variable and standard errors clustered at the state level. In all regressions, the independent variable of interest is the time varying marginalization index for the 1km radius around the facility. Facility-level fixed effects and time fixed effects are estimated but not presented. ** $p<0.05$, * $p<0.10$ for one-sided tests. In Panel A, we run a first difference model with two time periods, an aggregated early period representing 2005-2007 and an aggregated late period representing 2011-2013. In Panel B, we run a standard facility-by-year panel model with facility-level fixed effects. In Panel C, we include municipality interacted with year fixed effects to account for time varying factors within municipalities in particular sorting within municipalities based on annual changes in pollution.

Table 2. Renter shares and voter turnout as moderators of the pollution-marginalization relationship

Panel A: Interactions w/ Share of Housing Renter Occupied. First differences, facility by period data.							
DEP.VAR: The log of:	As	Cd	Cr	CN-	Pb	Hg	Ni
Marginalization Index	0.473**	0.576**	0.488**	0.685**	0.566**	1.058**	0.748**
	(0.132)	(0.261)	(0.233)	(0.190)	(0.233)	(0.268)	(0.228)
Marginalization ×	-0.017**	-0.020^{+}	-0.011	-0.020**	-0.013*	-0.036**	-0.025**
Renter share	(0.006)	(0.012)	(0.010)	(0.009)	(0.007)	(0.013)	(0.008)
# observations	771	752	749	776	839	710	861
Panel B: Interactions w/ Share of Housing Renter Occupied. Full panel, facility by year data.							
DEP.VAR: The log of:	As	Cd	Cr	CN-	Pb	Hg	Ni
Marginalization Index	0.368**	0.179	0.319^{+}	0.459**	0.310^{+}	0.612**	0.544**
	(0.164)	(0.224)	(0.211)	(0.155)	(0.233)	(0.230)	(0.178)
Marginalization ×	-0.016*	0.004	-0.005	-0.011^{+}	-0.008	-0.018^{+}	-0.015^{+}
Renter share	(0.009)	(0.009)	(0.011)	(0.008)	(0.011)	(0.011)	(0.009)
# observations	4,211	4,079	4,107	4,236	4,431	3,995	4,438
Panel C: Interactions w/ Voter Turnout. First differences, facility by period data.							
DEP.VAR: The log of:	As	Cd	Cr	CN-	Pb	Hg	Ni
Marginalization Index	0.701	0.463	-0.255	0.082	-0.458	1.058^{+}	0.564
	(0.688)	(0.945)	(0.581)	(0.788)	(1.129)	(0.679)	(0.529)
Marginalization ×	-0.009	-0.004	0.009	0.004	0.013	-0.011	-0.005
Voter Turnout	(0.011)	(0.015)	(0.010)	(0.013)	(0.018)	(0.010)	(0.009)
# observations	775	754	751	775	839	712	865
Panel D: Interactions w/ Voter Turnout. Full panel, facility by year data.							
DEP.VAR: The log of:	As	Cd	Cr	CN-	Pb	Hg	Ni
Marginalization Index	0.933	0.675	-0.048	0.454	-0.059	1.708**	1.018**
	(0.911)	(0.584)	(1.002)	(0.751)	(0.762)	(0.740)	(0.485)
Marginalization ×	-0.013	-0.007	0.005	-0.003	0.004	-0.023*	-0.012^{+}
Voter Turnout	(0.015)	(0.009)	(0.016)	(0.012)	(0.012)	(0.012)	(0.008)
# observations	4,221	4,087	4,122	4,241	4,436	3,999	4,456

Notes: Regressions include population density (time-varying), facility FEs, and time FEs. Standard errors clustered at the state level. ** $p<0.05$, * $p<0.10$, $^{+}$ $p<0.20$ for two-sided tests. The coefficient on marginalization (uninteracted) represents the effect of marginalization holding the interacted socio-demographic measure at 0, which may or may not have economic or practical content (e.g., the supports of our renter share and vote share don't approach 0).

Table 3. Sorting: Relationships between pollution and population turnover

Panel A: OLS Regressions of long-term residents on pollution

DEP.VAR:	% of local residents residing in the municipality for at least 5 years						
	(As)	(Cd)	(Cr)	(CN-)	(Pb)	(Hg)	(Ni)
Pollution (levels)	0.006	-0.000	-0.001	-0.005	-0.000	-0.007	-0.001
	(0.009)	(0.003)	(0.001)	(0.006)	(0.001)	(0.033)	(0.001)
# observations	1,428	1,413	1,424	1,439	1,478	1,414	1,478

Panel B: OLS Regressions of long-term residents on pollution and marginalization

DEP.VAR:	% of local residents residing in the municipality for at least 5 years						
	(As)	(Cd)	(Cr)	(CN-)	(Pb)	(Hg)	(Ni)
Pollution (levels)	0.003	-0.001	-0.001	-0.006	-0.000	-0.011	-0.001
	(0.009)	(0.003)	(0.001)	(0.006)	(0.001)	(0.034)	(0.001)
Marginalization Index	2.033**	2.098**	2.056**	2.072**	1.994**	1.952**	2.084**
	(0.819)	(0.830)	(0.824)	(0.869)	(0.838)	(0.865)	(0.832)
# observations	1,428	1,413	1,424	1,439	1,478	1,414	1,478

Panel C: First differences regressions, facility by period data

DEP.VAR:	% of local residents residing in the municipality for at least 5 years						
	(As)	(Cd)	(Cr)	(CN-)	(Pb)	(Hg)	(Ni)
Pollution (logs)	0.030	-0.004	-0.090	-0.080	-0.097	-0.014	-0.067
	(0.056)	(0.040)	(0.058)	(0.053)	(0.075)	(0.047)	(0.047)
Population density	X	X	X	X	X	X	X
Year FE	X	X	X	X	X	X	X
Facility FE	X	X	X	X	X	X	X
# observations	1,004	958	980	1,042	1,052	946	1,058

NOTES: Each column reports coefficients from regressions with the share of local residents residing in the municipality for at least 5 years as the dependent variable and standard errors clustered at the state level. In all regressions, the independent variable of interest is the time varying log of pollution discharges. The difference across the columns is that log pollution applies to discharges of As, Cd, Cr, CN-, Pb, Hg, and Ni, respectively. In Panels A and B, a constant is included in each regression. ** $p<0.05$, * $p<0.10$ for two-sided tests. In Panels A and B, we regress cross-sectional long-run resident share averaged over all census periods for each facility on cross-sectional pollution averaged over all annual periods for each facility (and controls). In Panel C, we regress the change in long-run resident share between 2000 and 2010 for each facility on the change in pollution between a 2005-2007 aggregate period and a 2008-2010 aggregate period for each facility (and controls). We lack the data to replicate a full annual panel analysis with the share of local residents residing in the municipality for at least 5 years as the dependent variable.

APPENDIX:

Data Appendix
Appendix (Robustness) Tables

Lopamudra Chakraborti and Jay Shimshack

"Environmental disparities in urban Mexico:
Evidence from toxic water pollution"

DATA APPENDIX

Data Appendix Table. In-Sample vs. Out-of-Sample facilities

	Out of Sample Facilities	In Sample Facilities	Difference	p-value
INDUSTRY				
Automotive	0.10	0.10	-0.00	0.65
Chemicals	0.23	0.31	-0.08	0.00
Electronics	0.04	0.07	-0.03	0.00
Energy	0.04	0.01	0.03	0.00
Food	0.07	0.07	-0.01	0.48
Metals	0.17	0.16	0.01	0.42
Other	0.22	0.16	0.05	0.00
Petroleum	0.07	0.04	0.03	0.00
Concrete	0.04	0.04	-0.00	0.74
Wood	0.03	0.04	-0.01	0.40
REGION				
Center	0.27	0.39	-0.12	0.00
Northeast	0.32	0.31	0.00	0.83
Northwest	0.06	0.05	0.01	0.21
South	0.07	0.05	0.01	0.11
Southeast	0.10	0.02	0.07	0.00
West	0.19	0.17	0.02	0.18
POLLUTION (kg)				
Arsenic (As)	5.44	3.17	2.27	0.03
Cadmium (Cd)	13.46	9.45	4.01	0.18
Chromium (Cr)	30.23	30.11	0.13	0.99
Cyanide (CN-)	8.43	7.10	1.33	0.42
Lead (Pb)	39.42	35.80	3.62	0.69
Mercury (Hg)	1.42	0.74	0.68	0.02
Nickel (Ni)	58.81	38.86	19.95	0.08

NOTES: Values for in-sample facilities are the same as those in Table 1. Relative to non-sample facilities, sample facilities are more likely to be in the chemicals and electronics sectors and less likely to be in the energy and petroleum industries. Sample facilities are more likely to be in the central region and less likely to be in the Southeast region. Sample facilities have lower average discharges of As and Hg. However, average discharges of Cd, Cr, CN-, Pb, and Ni are statistically indistinguishable between sample and non-sample facilities.

Data Appendix Table. Cross-sectional Summary Statistics

	Facilities	Mean	Std. Dev.	Max.
INDUSTRY				
Automotive	1,631	0.10	0.31	1.00
Chemicals	1,631	0.31	0.46	1.00
Electronics	1,631	0.07	0.25	1.00
Energy	1,631	0.01	0.09	1.00
Food	1,631	0.07	0.26	1.00
Metals	1,631	0.16	0.37	1.00
Other	1,631	0.16	0.37	1.00
Petroleum	1,631	0.04	0.20	1.00
Concrete	1,631	0.04	0.19	1.00
Wood	1,631	0.04	0.19	1.00
REGION				
Center	1,631	0.39	0.49	1.00
Northeast	1,631	0.31	0.46	1.00
Northwest	1,631	0.05	0.22	1.00
South	1,631	0.05	0.22	1.00
Southeast	1,631	0.02	0.15	1.00
West	1,631	0.17	0.38	1.00
SOCIO-DEMOGRAPHICS				
Marginalization Index	1,631	-0.76	0.87	3.50
% not attending school	1,631	4.12	1.94	35.63
% no access to healthcare	1,631	34.75	11.09	75.87
% without sanitary drainage	1,631	6.01	10.36	96.97
% without piped water	1,631	19.92	19.49	100.00
% without refrigerator	1,631	13.11	11.31	84.78
% in overcrowded housing	1,631	21.68	10.76	71.81
% indigenous language	1,625	1.21	1.66	30.87
% renter occupied housing	1,542	19.72	11.05	100.00
% local municipality residents	1,620	89.72	5.97	97.98
Child mortality rate	1,631	3.28	1.16	10.57
Population Density	1,631	7.11	5.17	28.70
POLLUTION (kg)				
Arsenic (As)	1,439	3.17	20.31	502.00
Cadmium (Cd)	1,424	9.45	65.59	1,445.21
Chromium (Cr)	1,435	30.11	199.45	4,305.57
Cyanide (CN-)	1,450	7.10	46.60	883.01
Lead (Pb)	1,489	35.80	253.15	4,540.00
Mercury (Hg)	1,425	0.74	5.10	119.29
Nickel (Ni)	1,489	38.86	221.51	4,653.54

NOTES: Industry and location values are time invariant. To create the cross section for the other variables, we collapse by taking means for *i* over all *t* with non-missing values. For socio-demographics, this typically involves averaging over the 2000, 2005, and 2010 censuses. For pollution, this involves averaging over annual data 2005-2013. Not every facility pollutes all 7 toxic pollutants. As consequence, the number of reporting facilities for any given contaminant in the Table does not equal the full 1,631 sample size.

ADDITIONAL DATA NOTES

Plant locations

The industrial facilities in our sample are located in highly urbanized areas. Even after restricting the analysis to only urban AGEBs, comparisons of plant locations suggest that areas within a 1km radius of our sample facilities have higher population density (roughly 50% more persons per square km), lower marginalization, and less racial variability than the less urbanized areas more than 1km from sample facilities. We present summary statistics in the Table below.

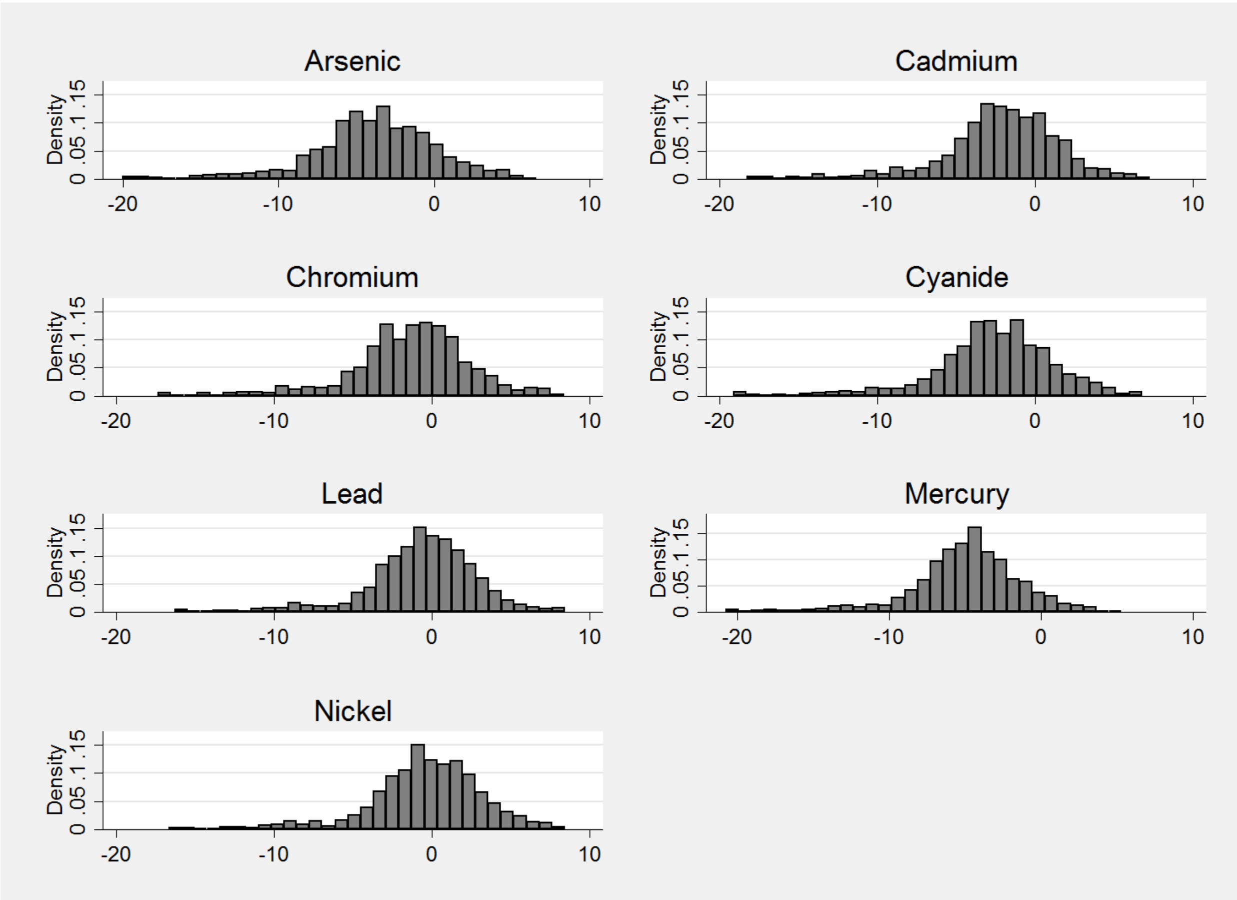
Data Appendix Table. Plant locations vs. non-plant locations, urban AGEBs only

2010 demographics	Within 1km of a sample plant	Outside 1km of a sample plant
Marginalization Index	-.59	.02
Population Density	9,362	6,188
Race	1.24	5.22

Additional graphs and summary statistics for pollution

As the figure below illustrates, the underlying pollution distributions are restricted to the positive domain and heavily right skewed

Data Appendix Figure. Probability Distribution Functions, by pollutant, in logs.

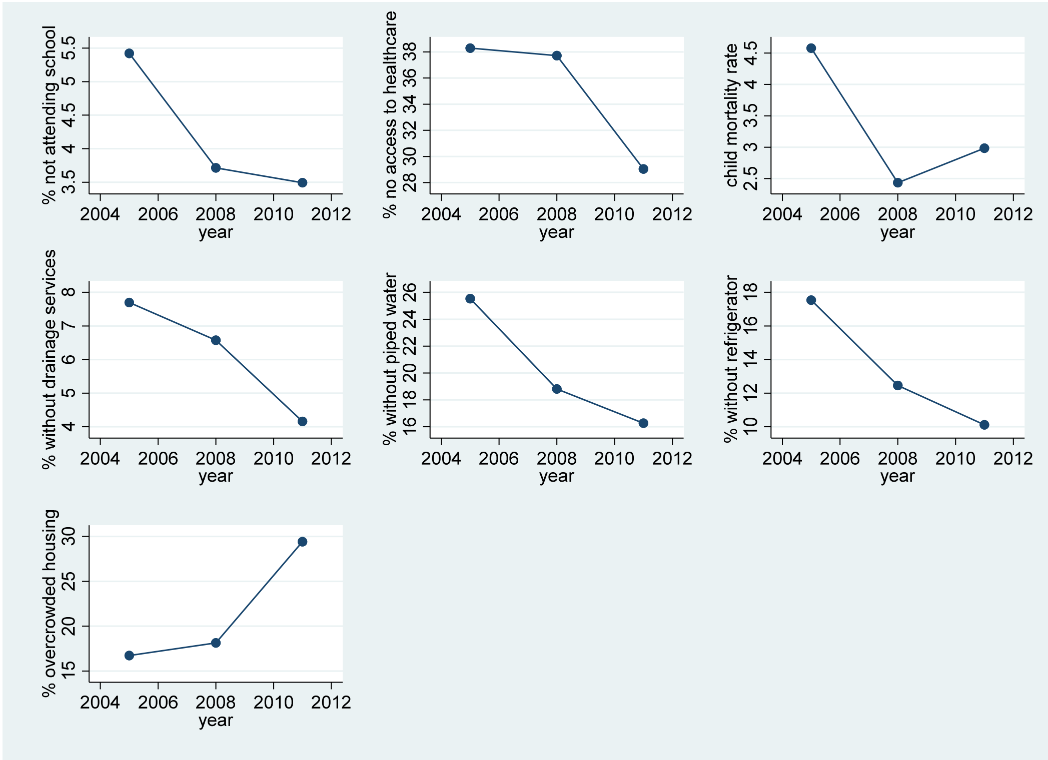


Notes. Untransformed pollution is restricted to the positive domain and heavily skewed to the right. After logging pollution discharges, distributions are approximately normal. Means slightly less than one correspond to median untransformed pollution discharges slightly less than 1kg.

Marginalization Index Components

Variables making up the marginalization index, depending on the census year, may include: population earning less than 2x minimum wage; population 6-14 years old not attending school; population over 15 without post-primary education; population over 15 without secondary education; household without piped water; household without septic connection; household without adequate drainage; household with mud floor; household without refrigerator; household without adequate roofing; household with overcrowding; child mortality rates; population without access to basic health services; teenage births; and other measures. The individual components determining the score may vary from year to year. In principle, this means that the index always "ranks" local area marginalization appropriately in any given time period but the index values may not be fully comparable across time. In practice, this technical detail is not important for our analysis. Results are robust to using raw CONAPO marginalization indexes or our own marginalization "scores" calculated using CONAPO methods applied to the individual census questions. Results are robust to using individual wealth measures rather than a summary index. Our panel analyses include time fixed effects which net out average differences common to all cross-sectional units and implicitly rescale the marginalization index for each period.

Data Appendix Figure. Trends in the components of the marginalization index over time

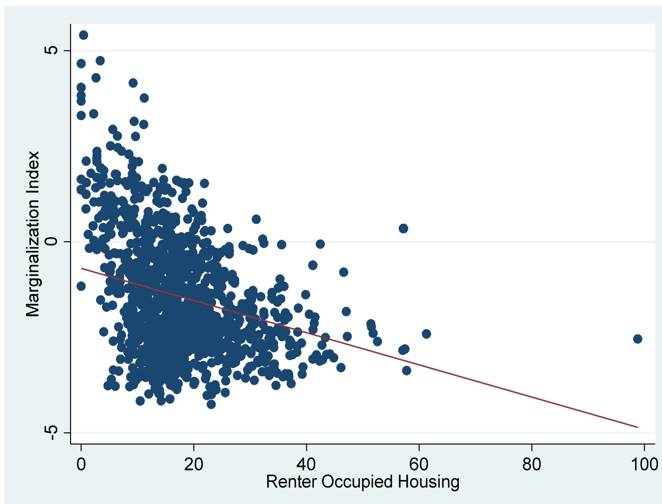


As noted in the figure, most components of marginalization are decreasing (improving) over time, although non-monotonically, including: population ages 6-14 not attending school; households without piped water; households without a refrigerator; mortality rates; and population without access to basic health services. In contrast, households with overcrowding are increasing over time in our sample.

Marginalization and Renter Occupied Housing

Empirically, in our sample, marginalization is negatively correlated with rental rates. As noted in the figure below, the correlation coefficient between marginalization and the share of renter occupied housing is -0.27. The figure documents that locations with very high marginalization have very low shares of renter occupied housing and that areas with above median shares of renter occupied housing are associated with low marginalization on average.

Data Appendix Figure. **Scatterplot of marginalization vs. the share of renter housing**



Defining administrative locations

We use latitude and longitude to establish administrative locations like AGEB (of >50,000), municipality (of >2,000), state (of 32), and region (of 6). In the 2000s, Mexico had ~2350 municipalities with an average size of ~800 square kilometers and a population of ~47,000 people. Mexico had 32 states (including Distrito Federal, or Mexico City) with an average size of ~60,000 square kilometers and a population of 3.5 million people. We also aggregate Mexican states into 6 regions: northwest (Baja California, Baja California Sur, Sinaloa, Sonora); west (Aguascalientes, Colima, Guanajuato, Jalisco, Michoacan, Nayarit, Queretaro); center (Distrito Federal, Mexico, Hidalgo, Morelos, Puebla, Tlaxcala); south (Guerrero, Oaxaca, Chiapas, Veracruz); Southeast (Campeche, Quintana Roo, Tabasco, Yucatan); and Northeast (Coahuila, Nuevo Leon, Tamaulipas, Chihuahua, Durango, Zacatecas, and San Luis Potosi).

MISSING DATA

RETC pollution data do not comprise a "square" dataset, and many facility-by-year pollution values are missing. Roughly 20% of missing values are attributable to plants that never report on a given contaminant for the entire sample. For example, a facility in our dataset may report on Ar, Cd, Cr, Pb, and Ni emissions for one or more years but never report CN- and Hg emissions. The other 80% of missing values are unexplained in the RETC database. These values may represent: (a) data missing at random due to reporting or data administration issues, (b) zero discharges or discharges below the required RETC reporting threshold in a given year for that contaminant, or (c) data missing strategically to hide undesirable behavior. We have no way of knowing which of these explanations applies.

We address missingness as follows. In our analysis, we generally assume explanation (a), noting that random measurement error results in attenuation bias. We also explore robustness to explanation (b) by replicating all analyses with missing values set to zero discharges below. Regarding missing data explanation (c), we take both an institutional and empirical perspective. First, we note that (c) may be less likely than imagined. RETC is not a regulatory database. Personnel from Mexico's environment ministry assess recordkeeping and pollution reporting as part of their air, water, and waste programs, but reporting unusually high toxic water pollution discharges into RETC cannot and does not result in direct government penalties, fines, or sanctions. Second, we recognized that bias will arise in our analysis if "missingness" appears differentially across poor vs. rich neighborhoods (or high majority race vs. high minority race neighborhoods) and considered the empirical implications. We therefore attempted to predict "missingness" econometrically. We constructed square panels for each contaminant and regressed a 0/1 missingness indicator on the marginalization index, our indigenous race measure, industry fixed effects, state fixed effects, and year fixed effects. We present frequency of missing data for each pollutant by year in *Missing Data Appendix Table 1*.

Missing Data Appendix Table 1. Frequency of Missingness, by pollutant and year

Year	As	Cd	Cr	CN-	Pb	Hg	Ni
2005	1029	1036	1031	1015	988	1049	996
2006	923	938	947	914	873	943	894
2007	1016	1029	1047	1004	974	1034	981
2008	1267	1283	1267	1249	1233	1287	1242
2009	1128	1176	1169	1128	1134	1162	1137
2010	1238	1263	1254	1248	1236	1283	1226
2011	1260	1268	1257	1269	1242	1284	1226
2012	1193	1201	1197	1199	1168	1218	1152
2013	1224	1231	1224	1233	1215	1250	1191

NOTES: Total of 1,631 facilities with one of the seven toxics reported at least once 2005-2013.

Missing Data Appendix Table 2 indicates that missingness, although not purely random, is reassuringly not systematically related to sociodemographics. Missingness is increasing over time (although non-monotonically) and more likely in some states than others (variation across states differs by substance). Conditional on other facility characteristics, however, missingness is not typically associated with wealth or race.

Missing Data Appendix Table 2. Predictions of Missingness, by pollutant

Regressions of a binary variable for missing pollution data on socio-demographics and controls

DEP.VAR: 0/1 indicator of:	As	Cd	Cr	CN-	Pb	Hg	Ni
Marginalization Index	0.005	-0.003	-0.009	-0.012*	-0.002	-0.002	-0.002
	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)
Indigenous race	-0.003	-0.004	-0.003	-0.002	-0.001	-0.002	-0.001
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Industry FE	X	X	X	X	X	X	X
State FE	X	X	X	X	X	X	X
Year FE	X	X	X	X	X	X	X
# observations	12,906	12,771	12,870	12,996	13,356	12,780	13,347

NOTES: Linear probability estimates. ** p<0.05, * p<0.10 for two-sided tests.

We also explored the possibility that missing pollution observations were truly zero (or very low) discharges. To operationalize tests, we replaced all missing observations with zeros provided the facility reported on the given pollutant at least once during our sample period. We then replicated all analyses. We find smaller but still robustly positive relationships between pollution and marginalization (see *Missing Data Appendix Table 3*). For example, full panel fixed effects models indicate that a 1 standard deviation increase in the marginalization index for the neighborhood around a facility results in: a 19.1% increase in arsenic, a 13.8% increase in cadmium, a 20.6% increase in chromium, a 20.8% increase in cyanides, a 11.7% increase in lead, a 28.0% increase in mercury, and a 14.4% increase in nickel.

Missing Data Appendix Table 3. Robustness to Replacing Missing with Zero

Panel A: Cross-sectional regressions of pollution on marginalization and population density

DEP.VAR:	As	Cd	Cr	CN-	Pb	Hg	Ni
Marginalization Index	0.015	1.596*	4.061*	0.358	2.465*	0.093**	1.123
	(0.233)	(1.184)	(2.771)	(0.352)	(1.560)	(0.043)	(3.084)
Population density	X	X	X	X	X	X	X
# observations	1,439	1,424	1,435	1,450	1,489	1,425	1,489

Panel B: Panel regressions: First differences, facility by period data

DEP.VAR: The log of:	As	Cd	Cr	CN-	Pb	Hg	Ni
Marginalization Index	0.245*	0.128	0.302**	0.246**	0.201**	0.303**	0.156**
	(0.156)	(0.110)	(0.124)	(0.132)	(0.083)	(0.179)	(0.084)
Population density	X	X	X	X	X	X	X
Facility FE	X	X	X	X	X	X	X
# observations	2,771	2,745	2,766	2,795	2,868	2,742	2,872

Panel C: Panel regressions: Full panel, facility by year data

DEP.VAR: The log of:	As	Cd	Cr	CN-	Pb	Hg	Ni
Marginalization Index	0.163*	0.118**	0.176**	0.178**	0.100**	0.239**	0.123**
	(0.099)	(0.050)	(0.051)	(0.063)	(0.036)	(0.106)	(0.033)
Population density	X	X	X	X	X	X	X
Year FE	X	X	X	X	X	X	X
Facility FE	X	X	X	X	X	X	X
# observations	12,480	12,363	12,456	12,591	12,924	12,351	12,936

Panel D: Full panel regressions, municipality with year interactions

DEP.VAR: The log of:	As	Cd	Cr	CN-	Pb	Hg	Ni
Marginalization Index	0.044	0.110*	0.134**	0.077	0.076	0.100	0.163**
	(0.099)	(0.078)	(0.077)	(0.070)	(0.079)	(0.095)	(0.095)
Population density	X	X	X	X	X	X	X
Municipality X Year FE	X	X	X	X	X	X	X
Facility FE	X	X	X	X	X	X	X
# observations	12,480	12,363	12,456	12,591	12,924	12,351	12,936

NOTES: All panels present results from regressions that mimic those reported Table 1 with missing pollution numbers replaced with zero. Standard errors clustered at the state level. ** p<0.05, * p<0.10 for one-sided tests.

Appendix (Robustness) Tables

Appendix Table 1. Robustness to 0.5 KM Radii

Panel A: Cross-sectional regressions of pollution on marginalization and population density							
DEP.VAR:	As	Cd	Cr	CN-	Pb	Hg	Ni
Marginalization Index	0.71	3.83**	9.19**	0.58	21.37**	0.16*	8.25
	(0.71)	(2.11)	(4.17)	(1.43)	(9.18)	(0.10)	(7.29)
Population density	X	X	X	X	X	X	X
# observations	1,315	1,303	1,315	1,330	1,366	1,304	1,363
Panel B: Panel regressions: First differences, facility by period data							
DEP.VAR: The log of:	As	Cd	Cr	CN-	Pb	Hg	Ni
Marginalization Index	0.09	0.09	0.12	0.25	0.25	0.29	0.28*
	(0.24)	(0.26)	(0.19)	(0.26)	(0.21)	(0.29)	(0.18)
Population density	X	X	X	X	X	X	X
Facility FE	X	X	X	X	X	X	X
# observations	712	679	683	713	760	662	785
Panel C: Panel regressions: Full panel, facility by year data							
DEP.VAR: The log of:	As	Cd	Cr	CN-	Pb	Hg	Ni
Marginalization Index	0.08	0.25	0.20	0.25*	0.18	0.35**	0.29**
	(0.18)	(0.20)	(0.19)	(0.16)	(0.15)	(0.18)	(0.15)
Population density	X	X	X	X	X	X	X
Year FE	X	X	X	X	X	X	X
Facility FE	X	X	X	X	X	X	X
# observations	3,798	3,662	3,717	3,819	3,983	3,616	4,006
Panel D: Full panel regressions, municipality with year interactions							
DEP.VAR: The log of:	As	Cd	Cr	CN-	Pb	Hg	Ni
Marginalization Index	0.23	0.51**	0.48**	0.41**	0.31*	0.50**	0.61**
	(0.26)	(0.24)	(0.24)	(0.22)	(0.19)	(0.20)	(0.21)
Population density	X	X	X	X	X	X	X
Municipality X Year FE	X	X	X	X	X	X	X
Facility FE	X	X	X	X	X	X	X
# observations	3,798	3,662	3,717	3,819	3,983	3,616	4,006

NOTES: All panels present results from regressions that mimic those reported Table 1. Standard errors clustered at the state level. ** p<0.05, * p<0.10 for one-sided tests. Average marginalization index and population density for the 500-meter radii are -0.57 and 7,996 people per square kilometer, respectively.

Appendix Table 2. Robustness to 1.5KM Radii

Panel A: Cross-sectional regressions of pollution on marginalization and population density							
DEP.VAR:	As	Cd	Cr	CN-	Pb	Hg	Ni
Marginalization Index	0.96	3.03**	13.62**	-0.37	15.60**	0.10	1.46
	(0.84)	(1.51)	(7.20)	(0.91)	(7.68)	(0.11)	(5.60)
Population density	X	X	X	X	X	X	X
# observations	1,439	1,424	1,435	1,450	1,489	1,425	1,489
Panel B: Panel regressions: First differences, facility by period data							
DEP.VAR: The log of:	As	Cd	Cr	CN-	Pb	Hg	Ni
Marginalization Index	0.07	-0.07	0.13	0.20*	0.16	0.16*	0.08
	(0.17)	(0.12)	(0.19)	(0.14)	(0.15)	(0.10)	(0.12)
Population density	X	X	X	X	X	X	X
Facility FE	X	X	X	X	X	X	X
# observations	789	768	768	793	856	726	882
Panel C: Panel regressions: Full panel, facility by year data							
DEP.VAR: The log of:	As	Cd	Cr	CN-	Pb	Hg	Ni
Marginalization Index	-0.02	0.12	0.10	0.17*	0.09	0.18*	0.17**
	(0.13)	(0.12)	(0.14)	(0.11)	(0.10)	(0.11)	(0.10)
Population density	X	X	X	X	X	X	X
Year FE	X	X	X	X	X	X	X
Facility FE	X	X	X	X	X	X	X
# observations	4,308	4,170	4,200	4,329	4,530	4,086	4,540
Panel D: Full panel regressions, municipality with year interactions							
DEP.VAR: The log of:	As	Cd	Cr	CN-	Pb	Hg	Ni
Marginalization Index	0.24	0.15	0.25	0.09	0.03	0.16	0.17
	(0.24)	(0.29)	(0.26)	(0.24)	(0.24)	(0.21)	(0.22)
Population density	X	X	X	X	X	X	X
Municipality X Year FE	X	X	X	X	X	X	X
Facility FE	X	X	X	X	X	X	X
# observations	4,308	4,170	4,200	4,329	4,530	4,086	4,540

NOTES: All panels present results from regressions that mimic those reported Table 1. Standard errors clustered at the state level. ** $p<0.05$, * $p<0.10$ for one-sided tests. Average marginalization index and average population density for the 1.5-kilometer radii are -0.6 and 10,250 people per square kilometer, respectively.

Appendix Table 3: Robustness to alternative lag structures

Panel A: Panel regressions: First differences, facility by period data							
DEP.VAR: The log of:	As	Cd	Cr	CN-	Pb	Hg	Ni
Marginalization Index	0.10	0.08	0.21	0.29**	0.29**	0.37**	0.21*
	(0.14)	(0.17)	(0.16)	(0.15)	(0.15)	(0.16)	(0.15)
Population density	X	X	X	X	X	X	X
Facility FE	X	X	X	X	X	X	X
# observations	773	752	752	778	841	712	864

Panel B: Panel regressions: Full panel, facility by year data							
DEP.VAR: The log of:	As	Cd	Cr	CN-	Pb	Hg	Ni
Marginalization Index	0.01	0.12	0.21	0.13	0.03	0.26**	0.21*
	(0.12)	(0.14)	(0.19)	(0.11)	(0.14)	(0.09)	(0.15)
Population density	X	X	X	X	X	X	X
Year FE	X	X	X	X	X	X	X
Facility FE	X	X	X	X	X	X	X
# observations	4,180	4,049	4,082	4,210	4,405	3,966	4,412

Panel C: Full panel regressions, municipality with year interactions							
DEP.VAR: The log of:	As	Cd	Cr	CN-	Pb	Hg	Ni
Marginalization Index	0.20	0.35	0.52*	0.23	0.18	0.54**	0.30
	(0.23)	(0.31)	(0.33)	(0.18)	(0.23)	(0.30)	(0.24)
Population density	X	X	X	X	X	X	X
Municipality X Year FE	X	X	X	X	X	X	X
Facility FE	X	X	X	X	X	X	X
# observations	4,180	4,049	4,082	4,210	4,405	3,966	4,412

Notes: Standard errors clustered at the state level. ** p<0.05, * p<0.10 for one-sided tests. Panel A replicates the first difference model where we assign 2005-2007 pollution data to census 2000 socioeconomic data (i.e. a 5-7 year lag) and 2011-2013 pollution data to census 2005 socioeconomic data (i.e. a roughly symmetric 6-8 year lag). Panels B and C replicates the panel models by similarly assigning pollution data from 2005-2009 to census 2000 (lags of 5-7 years) and assigning pollution data from 2010-2013 to census 2005 (roughly symmetric lags of 5-8 years).

Appendix Table 4. Replication with sociodemographic data from census 2000 and 2010

Panel A: Cross-sectional regressions of pollution on marginalization and population density

DEP.VAR:	As	Cd	Cr	CN-	Pb	Hg	Ni
Marginalization Index	0.96	2.92**	9.63*	0.45	13.93**	0.26*	2.83
	(0.76)	(1.46)	(5.83)	(0.80)	(5.58)	(0.17)	(4.52)
Population density	X	X	X	X	X	X	X
# observations	1,435	1,420	1,431	1,446	1,484	1,420	1,484

Panel B: Panel regressions: First differences, facility by period data

DEP.VAR: The log of:	As	Cd	Cr	CN-	Pb	Hg	Ni
Marginalization Index	0.18	0.20*	0.28*	0.32**	0.32**	0.38**	0.29**
	(0.14)	(0.15)	(0.17)	(0.14)	(0.17)	(0.15)	(0.14)
Population density	X	X	X	X	X	X	X
Facility FE	X	X	X	X	X	X	X
# observations	779	758	757	783	845	716	869

Panel C: Panel regressions: Full panel, facility by year data

DEP.VAR: The log of:	As	Cd	Cr	CN-	Pb	Hg	Ni
Marginalization Index	0.10	0.18*	0.18	0.11*	0.01	0.28**	0.17
	(0.11)	(0.14)	(0.16)	(0.08)	(0.12)	(0.08)	(0.13)
Population density	X	X	X	X	X	X	X
Year FE	X	X	X	X	X	X	X
Facility FE	X	X	X	X	X	X	X
# observations	4,198	4,071	4,102	4,229	4,420	3,987	4,434

Panel D: Full panel regressions, municipality with year interactions

DEP.VAR: The log of:	As	Cd	Cr	CN-	Pb	Hg	Ni
Marginalization Index	0.33**	0.38*	0.39	0.17	0.01	0.41*	0.20
	(0.17)	(0.25)	(0.33)	(0.14)	(0.19)	(0.29)	(0.17)
Population density	X	X	X	X	X	X	X
Municipality X Year FE	X	X	X	X	X	X	X
Facility FE	X	X	X	X	X	X	X
# observations	4,198	4,071	4,102	4,229	4,420	3,987	4,434

NOTES: All panels present results from regressions that mimic those reported Table 1 but utilizing only census 2000 and 2010 socioeconomic data. Standard errors clustered at the state level. ** p<0.05, * p<0.10 for one-sided tests.

Appendix Table 5. Replication with imputed sociodemographic data

Panel A: Panel regressions: Full panel, facility by year data							
DEP.VAR: The log of:	As	Cd	Cr	CN-	Pb	Hg	Ni
Marginalization Index	0.06 (0.17)	0.22 (0.20)	0.21 (0.22)	0.23* (0.15)	0.23* (0.17)	0.31** (0.16)	0.24* (0.16)
Population density	X	X	X	X	X	X	X
Year FE	X	X	X	X	X	X	X
Facility FE	X	X	X	X	X	X	X
# observations	4,268	4,126	4,157	4,290	4,489	4,042	4,496

Panel B: Full panel regressions, municipality with year interactions							
DEP.VAR: The log of:	As	Cd	Cr	CN-	Pb	Hg	Ni
Marginalization Index	0.34* (0.25)	0.27 (0.39)	0.39 (0.43)	0.31 (0.30)	0.11 (0.33)	0.29 (0.30)	0.26 (0.36)
Population density	X	X	X	X	X	X	X
Municipality X Year FE	X	X	X	X	X	X	X
Facility FE	X	X	X	X	X	X	X
# observations	4,268	4,126	4,157	4,290	4,489	4,042	4,496

Panel C: Full panel regressions, state with year interactions							
DEP.VAR: The log of:	As	Cd	Cr	CN-	Pb	Hg	Ni
Marginalization Index	0.34* (0.22)	0.37* (0.27)	0.37* (0.28)	0.29* (0.19)	0.28 (0.23)	0.37** (0.19)	0.38* (0.28)
Population density	X	X	X	X	X	X	X
State X Year FE	X	X	X	X	X	X	X
Facility FE	X	X	X	X	X	X	X
# observations	4,268	4,126	4,157	4,290	4,489	4,042	4,496

NOTES: All panels present results from regressions that mimic those reported Table 1 with imputed socioeconomic data. Standard errors clustered at the state level. ** p<0.05, * p<0.10 for one-sided tests.

Appendix Table 6. Naïve regressions without facility-level fixed effects

Panel A: Full panel, facility by year data, with state and year fixed effects							
DEP.VAR: The log of:	As	Cd	Cr	CN-	Pb	Hg	Ni
Marginalization Index	-0.05	-0.01	0.13*	0.09	0.05	0.07	0.05
	(0.08)	(0.08)	(0.08)	(0.09)	(0.07)	(0.09)	(0.06)
Population density	X	X	X	X	X	X	X
Year FE	X	X	X	X	X	X	X
State FE	X	X	X	X	X	X	X
# observations	4,250	4,116	4,148	4,277	4,470	4,033	4,483

Panel B: Full panel, facility by year data, with state, year, and industry fixed effects							
DEP.VAR: The log of:	As	Cd	Cr	CN-	Pb	Hg	Ni
Marginalization Index	-0.06	0.00	0.12*	0.09	0.06	0.07	0.06
	(0.08)	(0.09)	(0.09)	(0.09)	(0.08)	(0.10)	(0.07)
Population density	X	X	X	X	X	X	X
Year FE	X	X	X	X	X	X	X
State FE	X	X	X	X	X	X	X
Industry FE	X	X	X	X	X	X	X
# observations	4,250	4,116	4,148	4,277	4,470	4,033	4,483

NOTES: All panels present results from regressions that mimic those reported Table 1 without facility fixed effects. Standard errors clustered at the state level. ** p<0.05, * p<0.10 for one-sided tests.

Appendix Table 7. Robustness to alternative fixed effect specifications

Panel A: Full Panel regressions: state by year fixed effects							
DEP.VAR: The log of:	As	Cd	Cr	CN-	Pb	Hg	Ni
Marginalization Index	0.26**	0.35**	0.39**	0.32**	0.27**	0.40**	0.40**
	(0.13)	(0.16)	(0.16)	(0.11)	(0.12)	(0.12)	(0.16)
Population density	X	X	X	X	X	X	X
State by Year FE	X	X	X	X	X	X	X
Facility FE	X	X	X	X	X	X	X
# observations	4,250	4,116	4,148	4,277	4,470	4,033	4,483

Panel B: Full Panel regressions: industry by year fixed effects							
DEP.VAR: The log of:	As	Cd	Cr	CN-	Pb	Hg	Ni
Marginalization Index	0.09	0.23*	0.23*	0.30**	0.15	0.32**	0.26**
	(0.11)	(0.15)	(0.16)	(0.10)	(0.12)	(0.11)	(0.12)
Population density	X	X	X	X	X	X	X
Industry by Year FE	X	X	X	X	X	X	X
Facility FE	X	X	X	X	X	X	X
# observations	4,250	4,116	4,148	4,277	4,470	4,033	4,483

NOTES: All panels present results from regressions that mimic those reported in Table 1. Standard errors clustered at the state level. ** p<0.05, * p<0.10 for one-sided tests.

Appendix Table 8: Panel regressions dropping 87 facilities with name or ownership changes

Panel A: Panel regressions: First differences, facility by period data							
DEP.VAR: The log of:	As	Cd	Cr	CN-	Pb	Hg	Ni
Marginalization Index	0.18	0.14	0.21	0.34**	0.23*	0.35**	0.19
	(0.17)	(0.19)	(0.21)	(0.14)	(0.17)	(0.17)	(0.16)
Population density	X	X	X	X	X	X	X
Facility FE	X	X	X	X	X	X	X
# observations	736	721	718	740	792	683	828
Panel B: Panel regressions: Full panel, facility by year data							
DEP.VAR: The log of:	As	Cd	Cr	CN-	Pb	Hg	Ni
Marginalization Index	0.14	0.24**	0.21*	0.30**	0.14	0.36**	0.25**
	(0.12)	(0.14)	(0.15)	(0.10)	(0.12)	(0.12)	(0.12)
Population density	X	X	X	X	X	X	X
Year FE	X	X	X	X	X	X	X
Facility FE	X	X	X	X	X	X	X
# observations	4,032	3,901	3,936	4,053	4,218	3,833	4,255
Panel C: Full panel regressions, municipality with year interactions							
DEP.VAR: The log of:	As	Cd	Cr	CN-	Pb	Hg	Ni
Marginalization Index	0.46**	0.44**	0.46**	0.51**	0.26*	0.55**	0.42**
	(0.19)	(0.20)	(0.26)	(0.14)	(0.16)	(0.18)	(0.21)
Population density	X	X	X	X	X	X	X
Municipality X Year FE	X	X	X	X	X	X	X
Facility FE	X	X	X	X	X	X	X
# observations	4,032	3,901	3,936	4,053	4,218	3,833	4,255

Notes: Standard errors clustered at the state level; ** $p<0.05$, * $p<0.10$ for one-sided tests.