Title

Investigation of association between environmental and socioeconomic factors and preterm birth in California

Journal

Environment International, 121(Pt 2)

ISSN

0160-4120

Authors

Huang, Hongtai
Woodruff, Tracey J
Baer, Rebecca J
et al.

Publication Date

2018-12-01

DOI

10.1016/j.envint.2018.07.027

Peer reviewed

HHS Public Access
Author manuscript
Environ Int. Author manuscript; available in PMC 2019 December 01.

Published in final edited form as:
Environ Int. 2018 December ; 121(Pt 2): 1066–1078. doi:10.1016/j.envint.2018.07.027.

Investigation of Association between Environmental and Socioeconomic Factors and Preterm Birth in California

Hongtai Huang, PhD[a,b,*], **Tracey J. Woodruff, PhD, MPH**[b], **Rebecca J. Baer, MPH**[c,d], **Komal Bangia, MPH**[e], **Laura M. August, MPH**[e], **Laura L. Jellife-Palowski, PhD**[d,f], **Amy M. Padula, PhD, MSc**[b,§], and **Marina Sirota, PhD**[a,g,§,*]

[a]Bakar Institute for Computational Health Sciences, University of California, San Francisco, California, USA

[b]Program on Reproductive Health and the Environment, Department of Obstetrics, Gynecology and Reproductive Sciences, University of California, San Francisco, California, USA

[c]Department of Pediatrics, University of California, San Diego, California, USA

[d]California Preterm Birth Initiative, University of California, San Francisco, California, USA

[e]Office of Environmental Health Hazard Assessment, California Environmental Protection Agency, Sacramento, California, USA

[f]Department of Epidemiology and Biostatistics, University of California, San Francisco, California, USA

[g]Department of Pediatrics, University of California, San Francisco, California, USA

Abstract

Background: Preterm birth (PTB)[1], defined as birth at gestational age less than 37 weeks, is a major public health concern. Infants born prematurely, comprising of about 10% of the US newborns, have elevated risks of neonatal mortality and a wide array of health problems. Although numerous clinical, genetic, environmental and socioeconomic factors have been implicated in PTB, very few studies investigate the impacts of multiple pollutants and social factors on PTB using large scale datasets.

[1]PTB: Preterm Birth

[*]Corresponding author Correspondence and requests for materials should be addressed to H. Huang, University of California, San Francisco, Dept. of OB/GYN & Reproductive Science, Program on Reproductive Health and the Environment, Institute for Computational Health Sciences, 550 16th. Street, San Francisco, CA 94158; Telephone: 415.514.0517; Hongtai.Huang@ucsf.edu), or to M. Sirota, University of California, San Francisco, Institute for Computational Health Sciences, Mission Hall, 1599 4th St., San Francisco, CA 94158; Telephone: 415.514.0520; Marina.Sirota@ucsf.edu.
[§]These authors contributed equally to this work.

Author contributions:
H.H., T.J.W, L.J.P., A.P. and M.S. conceptualized the study. H.H. developed the study design, analysis and interpretation of the data, and drafted the manuscript. R.B. provided the OSHPD data and analytical insights into the study. K.B. and L.M.A. provided data and information on the drinking water contaminants of CalEnviroScreen 3.0 data. A.P. and M.S. supervised the study. All authors read and approved the final manuscript.

Competing interests: The authors declare no competing financial interests.

Objectives: To evaluate association between environmental and socioeconomic factors and PTB in California

Methods: We linked the birth cohort file maintained by the California Office of Statewide Health Planning and Development from 2009–2012 years across 1.8 million births and the CalEnviroScreen 3.0 dataset from California Communities Environmental Health Screening Tool at the census tract level for 56 California counties. CalEnviroScreen contains 7 exposure and 5 environmental effects variables that constitute the *Pollution Burden* variable, and 5 socioeconomic variables. We evaluated relationships between environmental exposures and the risk of PTB using hierarchical clustering analyses and GIS-based visualization. We also used logistic regression to evaluate the relationship between specific pollutant and exposure indicators and PTB, accounted for socio-demographic determinants such as maternal race/ethnicity, maternal age, maternal education and payment of delivery costs.

Results: There exists geographic variability in PTB for groups of counties with similar environmental and social exposure profiles. We found an association between *Pollution Burden,* particulate matter ≤ 2.5 μm ($PM_{2.5}$), and *Drinking Water Scores* and PTB (adjusted odds ratios were 1.03 (95% Confidence Interval (CI): 1.01, 1.04), 1.03 (95% CI: 1.02,1.04), and 1.04 (95% CI: 1.03,1.05), respectively). Additional findings suggest that certain drinking water contaminants such as arsenic and nitrate are associated with PTB in California.

Conclusions: CalEnviroScreen data combined with birth records offer great opportunity for revealing novel exposures and evaluating cumulative exposures related to PTB by providing useful environmental and social information. Certain drinking water contaminants such as arsenic and nitrate are potentially associated with PTB in California and should be investigated further.

Keywords

Environmental exposure; Environmental disparities; Social stressors; Drinking water contaminants; Preterm birth; Cumulative Risk

1. Introduction

Preterm birth (PTB), defined as birth at gestational age less than 37 weeks, is a major public health concern. According to the National Center for Health Statistics, the U.S. PTB prevalence was 9.85% in 2016, meaning approximately 1 of every 10 births were preterm (Martin et al. 2017b). While the definitive etiology of PTB remains unclear (Romero et al. 2014), environmental contamination has been implicated as a potential cause (Behrman and Butler 2007; Giorgis-Allemand et al. 2017). The majority of previous investigations of potential environmental causes of PTB have focused on air pollution including particulate matter (PM) (Brauer et al. 2008; Hao et al. 2016; Malley et al. 2017; Ritz et al. 2007; Sagiv et al. 2005; Schifano et al. 2013). Two separate meta-analyses included 23 and 18 individual studies respectively and confirmed a positive association between PM and increased PTB risk (Li et al. 2017; Sun et al. 2015). However, findings from individual studies of this association varied. Other environmental chemical exposures such as phthalate (Ferguson et al. 2014; Meeker et al. 2009), lead (Cantonwine et al. 2010; Vigeh et al. 2011), elemental carbon (Rappazzo et al. 2015a), sulfate (Rappazzo et al. 2015a), drinking water contaminant (*e.g.* arsenic (Ahmad et al. 2001; Laine et al. 2015), nitrate (Stayner et al. 2017)) have also been linked to increased PTB risk.

Social factors play an important part in PTB risk, acting as confounders, effect modifiers or both (Anthopolos et al. 2014; Braveman et al. 2015; Kramer et al. 2009; Masho et al. 2017; Wheeler et al. 2018; Zeka et al. 2008). A recent meta-analysis based on 45 studies concluded there are pronounced racial/ethnic disparities concerning PTB risk wherein African-American women (PTB prevalence: 13%) have an increased risk of PTB compared with white women (PTB prevalence: 9%) (Martin et al. 2017a; Schaaf et al. 2013). Racial discrimination could be an etiologic pathway by which race modifies PTB risk given its relationship to psychological distress (Giurgescu et al. 2012; Wheeler et al. 2018). Socioeconomic condition is an additional important but complex factor associated with PTB. Lower PTB rates are associated with higher socioeconomic advantage (Braveman et al. 2015). On a community level, neighborhoods with long-term high poverty experience increased PTB risk compared to those with low poverty (Margerison-Zilko et al. 2015), and material area deprivation is also linked to PTB (Auger et al. 2012). PTB also has a long-term impact on individual lifetime socioeconomic status attainment (Heinonen et al. 2013), suggesting a cycle of risk for vulnerable populations -- potentially over generations.

California is racially and ethnically diverse with distinct socioeconomic disparities. Approximately 14% Asian, 6.5% African American, and 38.9% Hispanics live in California in 2016. California has a 15% poverty rate with a median household income of about $62,000. Regarding educational attainment, more than 68% of the population (age > 25 years old) has less than a Bachelor's degree and around 18% has less than a high school degree (https://www.census.gov/). Such a high level of diversity in California communities has been connected to disproportionate burdens of environmental pollution and health disparities (Conroy et al. 2018; Morello-Frosch et al. 2001; Sadd et al. 2011). Also, evidence shows that neighborhood socioeconomic factors aggravate the positive relationship between PM and PTB, suggesting a 'double jeopardy' of high PM_{10} (PM <10 microns in aerodynamic diameter) exposure and living in impoverished neighborhood in the San Joaquin Valley of California (Padula et al. 2014a). Here, 'double jeopardy' was referred to the combined effects of environmental hazard exposures and vulnerability of socio-economic disadvantaged communities (Behrman and Butler 2007; Institute of Medicine 1999; Morello-Frosch and Shenassa 2006). Also, for the state of California, $PM_{2.5}$ and its constituents such as ammonium, nitrate and bromine have been shown to be associated with an increased risk for PTB, with certain demographic groups including African Americans and Asians experiencing greater impacts than others (Basu et al. 2017).

While associations between environmental and social factors with PTB have been evaluated in certain counties in California (Padula et al. 2014b; Ritz et al. 2007; Wilhelm et al. 2011), very few state-wide analyses examined the linkage between PTB and both environmental and social stressors (Basu et al. 2017). A recently published study considered cumulative environmental exposures and PTB across the U.S. (Rappazzo et al. 2015b). They identified an overall association between Environmental Quality Index and PTB; however, the environmental data used were summarized at the county level and were less current (2000–2005). Additionally, individual pollutant exposures were not evaluated directly. No study that we know have investigated both air pollutants and drinking water contaminants along with social factors at both the summary and individual levels on such a large scale. In this study, we integrated a large publicly available environmental exposure dataset (the California Communities Environmental Health Screening Tool (CalEnviroScreen 3.0)), together with a California State birth cohort dataset including birth certificate and hospital discharge data to investigate the association between environmental and socioeconomic factors and PTB in California across 1.8 million births and over 80 measures of chemical exposures.

2. Methods

2.1 Data

In this work, we used the birth cohort file maintained by the Office of Statewide Health Planning and Development (OSHPD) and the CalEnviroScreen 3.0 dataset.

2.1.1 California Birth Cohort—The birth cohort file contains linked birth and death certificates from all California livebirths, as well as detailed information on maternal and infant characteristics, hospital discharge diagnoses, and procedures recorded as early as one year prior to delivery and as late as one year postdelivery. Data files provided diagnoses and procedure codes based on the International Classification of Diseases, 9th Revision, Clinical Modification (ICD-9). The hospital inpatient discharge dataset is maintained by OSHPD based on reports submitted by thousands of individual and licensed healthcare facilities (https://www.oshpd.ca.gov/HID/). We utilized 1.8 million singleton birth records from this database from year 2009 to 2012, which included around 6,500 unique census tract identifiers, and focused on the following variables: maternal demographics (*e.g.* race/ ethnicity, gender, age, place of birth), socio-economic factors (*e.g.* education, insurance coverage) and birth characteristics (*e.g.* term vs. preterm, gestational age).

Methods and protocols for the study were approved by the Committee for the Protection of Human Subjects within the Health and Human Services Agency of the State of California.

2.1.2 CalEnviroScreen—To address the issue of environmental justice, the Office of Environmental Health Hazard Assessment (OEHHA), in the California Environmental Protection Agency (CalEPA), developed the California Communities Environmental Health Screening Tool (Cushing et al. 2015) to incorporate both chemical and non-chemical stressors to identify communities with disproportionate environmental and social burdens, and provide insights for relevant public health and environmental policy making. CalEnviroScreen is a screening tool used to identify California communities exposed to multiple sources of contamination (Faust et al. 2017). This tool integrates 20 indicator variables representing pollution and population vulnerability (Figure 1) for all 8,035 census tracts across all 58 counties in California. There are two main categories of indicators: *Pollution Burden* and *Population Characteristics*. *Pollution Burden* is the average of its two component scores: *Pollution Exposures* and *Environmental Effects*, with the latter half-weighted. *Population Characteristics* is the average of its two components: *Sensitive Populations* and *Socioeconomic Factors* (Faust et al. 2017). We used the data available from CalEnviroScreen 3.0 for 15 individual pollutant indicators and 2 summary indicator variables (*Pollution Burden* and *Drinking Water Score*) at the census tract-level. We also analyzed 13 individual drinking water contaminants that comprise the *Drinking Water Score* indicator, which is one of the pollution exposure indicators. Due to low variability across birth records in the matched dataset, two drinking water contaminants *Hexavalent Chromium* and *Combined Radium 226, 228* were not included in further analysis. Supplementary Material Table S-1 provides definitions of all the variables utilized in this study. We used raw values of these variables rather than their corresponding percentiles included in the database.

In this study, we used individual socioeconomic information from the California Birth Cohort including maternal race/ethnicity, maternal age, maternal education and payment of delivery costs, as confounding factors, but considered the neighborhood-level socioeconomic data from the CalEnviroScreen 3.0 as exposure variables, similar to other environmental contaminant variables. We assume that the exposure period for each birth is the entire pregnancy.

There were several steps involved in estimating drinking water contaminant concentrations by census tract developed for CalEnviroScreen 3.0. First, about 80 percent of drinking water system boundaries called community water systems (CWS) were identified based upon established boundaries from the Water Boundary Tool (http://cehtp.org/water/). The remainder of the boundaries were approximated based on locations of sampling wells or treatment plants and the population served by the systems. People living outside of CWS were assumed to drink from unregulated small water systems or private wells. A 6 × 6 mile township grid was used to summarize ambient groundwater concentrations and assigned to these areas (https://nationalmap.gov/small_scale/a_plss.html). Second, contaminant concentrations and violation data were associated and averaged within each water system or township for the selected group of contaminants and violation measures. Third, weighted averages were used to convert the contaminant concentrations or violation scores by water system or township to census tract. The weight used in the calculation was the population served by each system or population living in each township. Finally, contaminant or violation averages by census tract were computed and then summed to create the *drinking water score* for each tract.

2.2 Data Harmonization and Matching

We matched the OSHPD birth records with the CalEnviroScreen 3.0 data based on census tracts. Figure 2 shows the detailed process.

2.2.1 Initial Matching Strategy—With the assumption that geographic information of a tract would not change if a tract ID remains the same, we merged the OSHPD data with CalEnviroScreen 3.0 data based on census tract information which resulted in only 728,257 unique records with 3,953 unique geographic identifiers. In this initial matching, we were only able to match 40% of the records, due to incomplete or missing tract information for certain records in the OSHPD data. Among the 0.7 million records, we have 49,923 records of preterm birth (gestational age < 37 weeks) and 678,334 term birth records.

2.2.2 Secondary Matching Strategy—Because a proportion of census tract IDs in the OSPHD data were recorded based on tract ID systems across different decades, we were not able to match all records during initial matching. The U.S. Census Bureau updates the boundary information of all census tracts every ten years and provides census tract relationship files (https://www.census.gov/geo/mapsdata/data/relationship.html) for mapping how the boundary of a census tract may be defined differently over different tract ID systems. There are several possible types of boundary changes for a census tract including consolidation, split and other change (Logan et al. 2014).

To match the OSPHD records that were based on ID systems from different years within the CalEnviroScreen 3.0 data, we interpolated environmental information for census tracts from 2010 ID system to either 2000 or 1990 ID systems with the CalEnviroScreen 3.0 data by using the interpolation method with area weights, *i.e.* the area weighted method. This method has been applied previously and been shown to yield appropriate estimates for social variables such as poverty level and race/ethnicity if the estimates are not absolute counts as discussed in a previous study (Logan et al. 2014). This approach was compared with another method – the combined area and population interpolation method. The results showed that discrepancies in the estimates of absolute counts were significantly greater than average estimates. Because the majority of the environmental and social variable across all census tracts in the CalEnviroScreen 3.0 data were population-based average estimates, we chose the area weighted method.

With the 2010 U.S. census tract relationship files and the area weighted algorithm (Logan et al. 2014), we added to the CalEnviroScreen 3.0 dataset by matching it to both the 1990 and 2000 Tract ID systems. Different ID systems share a large number of census tracts IDs even though the boundary of a census tract with the same ID label may be defined differently over time, so those tracts may have more than one set of environmental records. For these tracts, we used existing information in the CalEnviroScreen 3.0 data rather than the expanded data. But for those unique tract IDs that were not included in the 2010 ID system, we used the information created by the area weighted algorithm. Eventually, we matched these expanded CalEnviroScreen data with the OSHPD PTB dataset, which resulted in 74% of the records matched. The remaining 22% could not be matched using this approach due to missing labels and other discrepancies related to census tract ID.

2.2.3 ZIP Code Interpolation Method—If no census tract ID was available for a birth record, a ZIP Code matching a valid census 2010 ZIP Code Tabulation Area was used as the surrogates of the birth's geographic information (Morello-Frosch et al. 2010). Except Placer, San Benito and Santa Cruz, most of the counties with empty labels and zero matched records are rural counties, including: Alpine, Amador, Calaveras, Colusa, Inyo, Lake, Lassen, Mendocino, Modoc, Mono, Plumas, Shasta, Sierra, Siskiyou, Tehama and Trinity. By applying the ZIP Code interpolation algorithms on the updated matched data, we were able to match 1,797,284 records (99.18%) in total.

2.3 Statewide Data Analysis and Visualization

To understand the dependence among the exposure variables, we examined and visualized the Pearson correlations between the environmental and social variables measured by the CalEnviroScreen 3.0 of the matched dataset by using the R package ‘corrplot’ (Wei and Simko 2017).

To visualize pollution levels and social factors together with birth data at the county level, we calculated the county-averaged contaminant exposure levels based on census tract-level data and generated a heatmap of the exposure levels normalized by the corresponding maximum of each variable. The structure of the heatmap was based on hierarchical clustering analysis embedded in the R function ‘heatmap.3’ (https://raw.githubusercontent.com/obigriffith/biostartutorials/master/Heatmaps/heatmap.3.R). According to the urban-rural scheme for counties developed by the CDC National Center for Health Statistics (NCHS), each California county was assigned with a discrete urbanization code (https://www.cdc.gov/nchs/data_access/urban_rural.htm) ranging from 1 (the highest level of urbanization) to 6 (the lowest level of urbanization). To identify the significant clusters, we calculated the statistical significance level for the county-wise hierarchical clustering using the R package ‘pvclust’ (Suzuki and Shimodaira 2006) that provides the Approximately Unbiased (AU) *p*-values based on multiscale bootstrap resampling scheme. In this study, the number of bootstrapping was set to be 1,000 times to gain reliable estimates. Clusters with AU *p*-values smaller than 0.05 were highlighted in blue rectangles indicating strong support by the data.

2.4 Association Analysis

We used logistic regression to evaluate the relationship between environmental pollutant data, social factors and PTB. Specifically, we used the R package ‘biglm’ (Lumley 2015) because of its ability to handle large scale database. We conducted two evaluations. First, we estimated the relationship between summary variables, *Pollution Burden* and *Drinking Water Score*, and PTB while accounting for potential confounders including maternal race/ethnicity, maternal age, maternal education and payment of delivery costs (private insurance or not). Second, we used multivariate logistic regression models by including all individual pollutant variables adjusted by all the above-mentioned confounders. For some pollutant variables identified as significantly associated with PTB, we also calculated and visualized their estimate arithmetic means for both the PTB and term birth groups for each county in California to evaluate their differences at the county level.

2.5 ArcGIS Mapping

Using ArcMap in ArcGIS for desktop (version 10.4; Esri Inc., Redlands, CA), we generated maps for contaminant variables of interest to evaluate the geographic variability of certain environmental pollutants across California.

3. Results

3.1 Study Population

The population in this study is ethnically diverse. More than 48% of births were Hispanic mothers, while non-Hispanic white, Asian and Black mothers accounted for approximately 25%, 14% and 6% of births respectively (Table 1). The majority of the women in the study (79.11%) were between 18 and 34 years at delivery. About half of the population had less than a college degree and a quarter with less than high school education at the time of delivery. Individuals with private insurance coverage constituted 45.84% of the population.

3.2 Statewide Data Analysis and Visualization Identifies Relationships Between Exposure Variables and Geographic Regions

To understand any potential dependence among the exposure variables from CalEnviroScreen 3.0, we observed and visualized correlations among different pairs of variables. Average Trihalomethane (THM) levels were negatively correlated with average nitrate level, and several social variables, including *Unemployment*, *Linguistic Isolation*, *Housing Burden* and *Poverty*, were positively correlated with each other (the upper left cluster in Figure 3). Another small cluster of positive correlation among variables *Solid Waste*, *Groundwater Threats*, *Cleanup Sites* and *Hazardous Waste* were observed (lower right of Figure 3). In addition, PCE was positively correlated with TCE (Figure 3).

There were several major groups of counties clustered together in the heatmap (Figures 4 & Supplementary Material Figure S-1). Cluster A (Figure 4, the leftmost, AU *p*-value = 0.02) of rural counties (Humboldt, Mendocino, Del Norte, Trinity, Tuolumne, Lassen, Siskiyou, see Figure 4, colored in light gray and gray in the second horizontal bar) with the lowest PTB prevalence is characterized by low level of all pollutants except 1,2-Dibromo-3-chloropropane average (DBCP). In contrast, cluster B (center to the left in Figure 4, AU *p*-value = 0.01) of suburban counties in the San Joaquin Valley (Kings, Tulane, Kern, Madera, Fresno & Stanislaus, colored orange in the second horizontal bar of Figure 4) are clustered together with PTB prevalence well above the average (indicated as deep purple color in the horizontal side bar at the top of Figure 4). These counties are characterized by high levels of social stressors such as poverty and environmental pollutants including ozone, $PM_{2.5}$, and arsenic and nitrate in drinking water. Cluster C (San Francisco, San Diego, Santa Clara, San Mateo, Marin, Sonoma, Sacramento, Alameda, Contra Costa, AU *p*-value = 0.02), centered in the middle of the heatmap shows modest level of PTB prevalence with several urban counties (Figure 4, colored in brown and black in the second horizontal bar). Other clusters have less strong exposure signals in comparison to these three clusters described. Majority of drinking water contaminants exhibited low variability (Figure 4, colored in blue). We found that AU *p*-values of Clusters A, B and C were all below 0.05 (Supplementary Material Figure S-1), which suggests high level of statistical significance regarding similarity of the exposure profiles of counties within the same cluster.

We observed two statistically significant clusters among the environmental variables along the vertical sidebar (Figures 4, Supplementary Material Figure S-2) when applying hierarchical clustering. *MCL Violation* and *TCR Violations* clustered close to the bottom

(AU *p*-value = 0.05), followed by three of the social variables, *Poverty, Unemployment* and *Housing Burden,* along with *Ozone* clustered at the bottom (AU *p*-value = 0.02).

3.3 Logistic Regression Analysis Identifies Specific Exposure Association with Preterm Birth

Due to a specific requirement of the R package 'biglm', we removed 348,684 birth records that contained missing information for the exposure variables of interest, so the sample size of the input dataset for regression analysis was 1,448,600 (Figure 2).

We found a positive association between *Pollution Burden* and PTB in our logistic regression. For an increase of 17.29 points (interquartile range (IQR) change) in *Pollution Burden* score, the adjusted odds ratio (AOR) was 1.03 (95% CI: 1.02,1.04, *p*-value < 0.0001) for PTB vs. term birth. The *Drinking Water Score* was also found to have a positive association with preterm birth – for an increase of 392.05 point (IQR change) in *Drinking Water Score*, the AOR was 1.04 (95% CI: 1.03,1.05, *p*-value < 0.0001) for PTB vs. term birth.

Among the individual indicator variables, we found that $PM_{2.5}$ and drinking water contaminants including nitrate and arsenic were positively associated with PTB (Table 2). Specifically, for an increase of 9.33 parts per million (ppm) (IQR change) in nitrate (as NO_3) average, the AOR was 1.02 (95% CI: 1.01, 1.03) for PTB vs. term birth; for an increase of 1.38 parts per billion (ppb) (IQR change) in arsenic average, the AOR was 1.01 (95% CI: 1.00, 1.01[2]) for PTB vs. term birth. However, THM in drinking water showed a negative association with PTB. Social factors including *Linguistic Isolation*, *Poverty*, *and Unemployment* were also found to elevate the PTB risk (Table 2).

In addition, we observed that associations between drinking water contaminants and PTB were higher in urbanized counties than in rural counties (Supplementary Material Table S-2). All social stressor variables that were significantly associated with PTB in urban or suburban counties were not associated with PTB in rural counties (Supplementary Material Table S-2).

With regard to the spatial variability of arsenic and nitrate across California, we found that northern California and the southern coastal areas have the lowest arsenic levels (less than 1.16ppb) in drinking water (colored in green in Figure 5). Higher arsenic concentrations were observed in Death Valley National Park, in the Mojave Desert, in the Mojave National Preserve and in Joshua Tree National Park, ranging from 2.54 to 32.09 ppb. Some of these areas (colored in purple in Figure 5) exceeded that the California Maximum Contaminant Level (MCL) for arsenic in drinking water (10 ppb) (https://www.waterboards.ca.gov/drinking_water/certlic/drinkingwater/Arsenic.shtml).

Similar to the geographic distribution of arsenic average concentration in drinking water in this state, high nitrate levels (33.04 to 85.48 ppm) were observed in the San Joaquin Valley area (Figure 6) which was above the California MCL for nitrate in drinking water (45 ppm,

[2]The 95% C.I. upper bound overlaps the adjusted odds ratio due to decimal points rounding.

details here: https://www.waterboards.ca.gov/rwqcb3/water_issues/programs/ag_waivers/docs/resources4grow ers/nitrate_info%20guide_102913.pdf), but less so in Death Valley.

The contrast of spatial variability of nitrate is less stark compared with that of arsenic. Urban areas of the state as well as northern California exhibit lower nitrate exposures overall except the adjacent areas of Highway 5.

3.4 Arsenic & Nitrate Concentrations Differ Between Preterm and Term Birth Groups

Heterogeneity across different counties was observed for both arsenic and nitrate contaminants based on their respective arithmetic means (Supplementary Material Figures S-3, S-4). The arsenic levels in drinking water for the PTB group were not always necessarily higher than those for the term birth group. For example, the average arsenic concentration for PTB was slightly higher than that for term birth in Kings county, although the arsenic levels for both groups were already well above the estimates of other counties. The levels of uncertainty which were denoted by the width of 95% confidence interval varied tremendously for different counties, ranging from less than 1 (*e.g.* arsenic levels in Alameda) to more than 25 ppb (*e.g.* arsenic levels in Inyo).

4. Discussion

This study integrates large and wide-ranging data on environmental exposures and detailed birth and hospital records for all singleton births in California over a 4-year period. This is one of the largest studies to date to examine the relationship between PTB and multiple environmental and social stressors. Our results are consistent with previous studies that have shown an association between PTB and exposures such as PM (Brauer et al. 2008; Hao et al. 2016; Ritz et al. 2007; Sagiv et al. 2005; Schifano et al. 2013). Our study also includes novel findings of relationships between PTB and exposures from drinking water contaminants including arsenic and nitrate. Additionally, our study included investigation of the relationship between PTB and social factors in conjunction with environmental factors and found that poverty and unemployment were also positively associated with PTB. These findings are in agreement with those of previous investigators (Auger et al. 2012; Margerison-Zilko et al. 2015; Rodrigues and Barros 2008). Lastly, it is important to note that this study contained individual birth records that were useful in accounting for potential confounders which improved the strength of the findings.

From our clustering analysis, we identified different clusters of counties based on the environmental and social factors investigated in this study. These clusters were also aligned with the distinct levels of both county urbanization and PTB prevalence. We found that PTB was associated with environmental pollutants provided that urbanization level of a county was intermediate or high. In rural areas, since fewer contaminant sites are located with smaller population, effects of pollutants along with social stressors are less impactful. In addition, several major clusters of counties had their own environmental or social ‘fingerprints’. That is to say, the most important variables linked to PTB differed across counties, which echoes the results of logistic regression models for each urbanization group.

Considering the results from logistic regression model, the most important environmental exposure variables positively related to PTB include $PM_{2.5}$, nitrate and arsenic. While statistically significant emergence of associations between $PM_{2.5}$ and PTB were expected, identification of associations between the other three drinking water contaminant variables and PTB were less so. The significant odds ratios identified may seem very small, but it may involve sizeable population impacts. For example, the OR associated with $PM_{2.5}$ is around 1.03, if the population without $PM_{2.5}$ exposure has a 7.0% PTB rate, then the sub-population with this exposure will have a 7.2% PTB rate. Such increase in PTB rate related to a specific pollutant exposure, though reflected as a small OR, may have public health policy implications.

Both nitrate and arsenic exhibited positive associations with PTB, but THM demonstrated the opposite. Maternal arsenic exposures have been previously linked to adverse birth outcomes (Ahmad et al. 2001; Laine et al. 2015). However, the linkage between arsenic and PTB is not fully clear (Myers et al. 2010). Our results suggest the potential role of arsenic exposure and PTB requires further study. Though associations between nitrate exposure in the form of herbicides and certain adverse pregnancy outcomes (*e.g.* small-for-gestational-age) have been identified (Ochoa-Acuna et al. 2009), the relationship between nitrate and PTB has not been fully evaluated (Ochoa-Acuna et al. 2009; Stayner et al. 2017). Previously, nitrate intake from public water supplies was found to be associated with significantly increased risk of thyroid disease (Ward et al. 2010), which could in turn lead to higher risk of PTB (Stagnaro-Green 2009). The positive association between nitrate and PTB identified in this study raised concerns and merit further investigation. The opposite trends of THM and other contaminants with PTB were possibly caused by the fact that THMs are disinfection by-products (https://www.cdc.gov/safewater/) in larger drinking water systems, especially in urban counties, but it could also be induced by confounding of other variables. The negative correlation between THM and nitrate was also captured in the correlation plot (Figure 3). Other drinking water contaminants such as PCE and TCE, which are large scale industrial contaminants emerging mostly in urban areas such as Los Angeles (shown in Figure 4) but not in the San Joaquin Valley area, were not observed to be significantly associated with PTB.

In addition, exposure variable Traffic was not found to be statistically associated with PTB across the whole state of California, although association between PTB and traffic-related air pollution has been identified previously (Laurent et al. 2016; Padula et al. 2014a). This may be related to the difference in the geographic exposure assignment involved in the variable Traffic across different studies. In the CalEnviroScreen 3.0 database, traffic density is defined as ‘vehicle-kilometers per hour per road length, within 150 meters of the census tract boundary’ (Supplementary Material Table S-1, more details in CalEnviroScreen 3.0 Report (Faust et al. 2017)), but previous studies oriented this variable based on maternal residence. For example, Laurent et al. estimated traffic density within ‘circular buffers of different sizes centered on maternal homes’ (Laurent et al. 2016), and Padula et al. calculated traffic density within ‘a 300m radius of geo-coded maternal residences (Padula et al. 2014a), instead of census tract boundaries. Such geographic difference involved in the definitions of traffic can contribute to the lack of associations we observed. In addition, the air pollution assessment methods in this study are quite different from those adopted by previous efforts such as timing and distance of measurements, which could potentially lead to different findings. We also suspect that some of the signals such as the association between $PM_{2.5}$ and PTB were 'diluted' in a statewide study comparing with studies focused on a particular area or region.

Strengths of the present study include large sample size, inclusion of a wide variety of exposure variables, integration of individual birth information to account for potential confounders, and useful visualization of the relationships between PTB and environmental and social factors while accounting for clustering of different counties. The CalEnviroScreen 3.0 data are more current and contain more accurate drinking water data in comparison to the study by Rappazzo *et al.*(Rappazzo et al. 2015b) Specifically, the drinking water contaminant data in CalEnviroScreen were based on chemical concentration monitoring at the census tract level in the California State Water Resources Control Board's Water Quality Monitoring database (https://www.waterboards.ca.gov/resources/data_databases/) while the water information involved in constructing the Environmental Quality Index is a collection of data from different sources at the county level (Messer et al. 2014). The higher resolution of drinking water contaminant data in this study allows us to pinpoint individual water pollutant variables that are associated with PTB, which could provide useful insights for developing relevant environmental and public health policy. There are also several limitations inherent in this study that should be recognized. First, the linkage of the OSHPD and CalEnviroScreen datasets could potentially result in imprecise effect estimates owing to misclassification of exposures. Second, we assumed that maternal residence was constant over the entire duration of pregnancy, which may not be valid across all 1.8 million birth records and could possibly introduce exposure misclassification into this study. Third, although we identified interesting environmental patterns associated with PTB, the CalEnviroScreen was created and designed for screening disadvantaged communities and therefore more suitable for community-based estimates, rather than to assign individual exposure levels. Due to the design of this tool/database, we assume the pollutant levels in this database to be constant during the entire pregnancy, and didn't account for temporal variability. Fourth, the methodology used to estimate drinking water contaminants involved population weighted averages by census tracts which may be covered by multiple drinking water systems and areas not served by public water systems, and relevant assumptions needed to be made due to inherent uncertainty. Therefore, some estimates may not be precise on the census-tract level, especially for rural counties. While these four limitations are certainly of concern, all of these items would likely pull our results towards the null and as such, we are less concerned about their leading to erroneous positive associations between contaminants and PTB. Finally, there is a broader limitation in that the results presented here are specific to the state of California. Broader national and international analyses are warranted to assess wider exposure-PTB relationships.

5. Conclusion

This comprehensive state-wide analysis compiles multiple pollutants and integrated social factors, which allowed us to explore associations between PTB and various contaminants that were not examined previously. The large sample size and diverse features of this study allowed us to better understand the linkage between environmental and social factors and

PTB on both the county and the state levels. These data could inform regulations to prevent PTB. We show that CalEnviroScreen data in combination with birth records offer great opportunity for revealing novel exposures and evaluating cumulative exposures related to PTB by providing useful environmental and social information, and that certain drinking water contaminants such as arsenic and nitrate are potentially associated with PTB in California and should be explored further. Small association signals may involve sizeable population impacts.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Funding:

This research was supported by funding from the California Preterm Birth Initiative at the University of California, San Francisco (UCSF7027075) and the National Institutes of Health (NIEHS R00ES021470, NLM K01LM012381). The funding sources had no involvement in the conduct of the research and preparation of the article.

References:

Ahmad SA, Sayed MH, Barua S, Khan MH, Faruquee MH, Jalil A, et al. 2001 Arsenic in drinking water and pregnancy outcomes. Environ Health Perspect 109:629–631. [PubMed: 11445518]

Anthopolos R, Kaufman JS, Messer LC, Miranda ML. 2014 Racial residential segregation and preterm birth: Built environment as a mediator. Epidemiology 25:397–405. [PubMed: 24681575]

Auger N, Park AL, Gamache P, Pampalon R, Daniel M. 2012 Weighing the contributions of material and social area deprivation to preterm birth. Soc Sci Med 75:1032–1037. [PubMed: 22698926]

Basu R, Pearson D, Ebisu K, Malig B. 2017 Association between pm2.5 and pm2.5 constituents and preterm delivery in california, 2000–2006. Paediatr Perinat Ep 31:424–434.

Behrman RE, Butler ASE. 2007 Preterm birth: Causes, consequences, and prevention. In: Preterm birth: Causes, consequences, and prevention Washington (DC):National Academies Press.

Brauer M, Lencar C, Tamburic L, Koehoorn M, Demers P, Karr C. 2008 A cohort study of traffic-related air pollution impacts on birth outcomes. Environ Health Persp 116:680–686.

Braveman PA, Heck K, Egerter S, Marchi KS, Dominguez TP, Cubbin C, et al. 2015 The role of socioeconomic factors in black-white disparities in preterm birth. Am J Public Health 105:694–702. [PubMed: 25211759]

Cantonwine D, Hu H, Sanchez BN, Lamadrid-Figueroa H, Smith D, Ettinger AS, et al. 2010 Critical windows of fetal lead exposure: Adverse impacts on length of gestation and risk of premature delivery. J Occup Environ Med 52:1106–1111. [PubMed: 21063188]

Conroy SM, Shariff-Marco S, Yang J, Hertz A, Cockburn M, Shvetsov YB, et al. 2018 Characterizing the neighborhood obesogenic environment in the multiethnic cohort: A multi-level infrastructure for cancer health disparities research. Cancer Causes Control 29:167–183. [PubMed: 29222610]

Cushing L, Faust J, August LM, Cendak R, Wieland W, Alexeeff G. 2015 Racial/ethnic disparities in cumulative environmental health impacts in california: Evidence from a statewide environmental justice screening tool (calenviroscreen 1.1). Am J Public Health 105:2341–2348. [PubMed: 26378826]

Faust J, August L, Bangia K, Galaviz V, Leichty J, Prasad S, et al. 2017 Calenviroscreen 3.0 report. California Environmental Protection Agency, Office of Environmental Health Hazard Assessment Sacramento, CA.

Ferguson KK, McElrath TF, Meeker JD. 2014 Environmental phthalate exposure and preterm birth. Jama Pediatr 168:61–67. [PubMed: 24247736]

Giorgis-Allemand L, Pedersen M, Bernard C, Aguilera I, Beelen RM, Chatzi L, et al. 2017 The influence of meteorological factors and atmospheric pollutants on the risk of preterm birth. Am J Epidemiol 185:247–258. [PubMed: 28087514]

Giurgescu C, Zenk SN, Dancy BL, Park CG, Dieber W, Block R. 2012 Relationships among neighborhood environment, racial discrimination, psychological distress, and preterm birth in african american women. J Obstet Gynecol Neonatal Nurs 41:E51–61.

Hao H, Chang HH, Holmes HA, Mulholland JA, Klein M, Darrow LA, et al. 2016 Air pollution and preterm birth in the u.S. State of georgia (2002–2006): Associations with concentrations of 11 ambient air pollutants estimated by combining community multiscale air quality model (cmaq) simulations with stationary monitor measurements. Environ Health Perspect 124:875–880. [PubMed: 26485731]

Heinonen K, Eriksson JG, Kajantie E, Pesonen AK, Barker DJ, Osmond C, et al. 2013 Latepreterm birth and lifetime socioeconomic attainments: The helsinki birth cohort study. Pediatrics 132:647–655. [PubMed: 24082003]

Institute of Medicine. 1999 In: Toward environmental justice: Research, education, and health policy needs Washington (DC).

Kramer MS, Lydon J, Seguin L, Goulet L, Kahn SR, McNamara H, et al. 2009 Stress pathways to spontaneous preterm birth: The role of stressors, psychological distress, and stress hormones. Am J Epidemiol 169:1319–1326. [PubMed: 19363098]

Laine JE, Bailey KA, Rubio-Andrade M, Olshan AF, Smeester L, Drobna Z, et al. 2015 Maternal arsenic exposure, arsenic methylation efficiency, and birth outcomes in the biomarkers of exposure to arsenic (bear) pregnancy cohort in mexico. Environ Health Perspect 123:186–192. [PubMed: 25325819]

Laurent O, Hu JL, Li LF, Kleeman MJ, Bartell SM, Cockburn M, et al. 2016 A statewide nested case-control study of preterm birth and air pollution by source and composition: California, 2001–2008. Environ Health Persp 124:1479–1486.

Li XY, Huang SQ, Jiao AQ, Yang XH, Yun JF, Wang YX, et al. 2017 Association between ambient fine particulate matter and preterm birth or term low birth weight: An updated systematic review and meta-analysis. Environ Pollut 227:596–605. [PubMed: 28457735]

Logan JR, Xu ZW, Stults BJ. 2014 Interpolating us decennial census tract data from as early as 1970 to 2010: A longitudinal tract database. Prof Geogr 66:412–420. [PubMed: 25140068]

Lumley T 2015 R package ‘biglm’ - bounded memory linear and generalized linear models

Malley CS, Kuylenstierna JC, Vallack HW, Henze DK, Blencowe H, Ashmore MR. 2017 Preterm birth associated with maternal fine particulate matter exposure: A global, regional and national assessment. Environ Int 101:173–182. [PubMed: 28196630]

Margerison-Zilko C, Cubbin C, Jun J, Marchi K, Fingar K, Braveman P. 2015 Beyond the crosssectional: Neighborhood poverty histories and preterm birth. Am J Public Health 105:1174–1180. [PubMed: 25880941]

Martin JA, Hamilton BE, Osterman MJ, Driscoll AK, Mathews TJ. 2017a Births: Final data for 2015. National vital statistics reports from the: Centers for Disease Control and Prevention, National Center for Health Statistics, National Vital Statistics System, 66(1),1.

Martin JA, Hamilton BE, Osterman MJK. 2017b Births in the united states, 2016. NCHS Data Brief:1–8.

Masho SW, Cha S, Chapman DA, Chelmow D. 2017 Understanding the role of violence as a social determinant of preterm birth. Am J Obstet Gynecol 216. [PubMed: 28412087]

Meeker JD, Hu H, Cantonwine DE, Lamadrid-Figueroa H, Calafat AM, Ettinger AS, et al. 2009 Urinary phthalate metabolites in relation to preterm birth in mexico city. Environ Health Perspect 117:1587–1592. [PubMed: 20019910]

Messer LC, Jagai JS, Rappazzo KM, Lobdell DT. 2014 Construction of an environmental quality index for public health research. Environ Health 13:39. [PubMed: 24886426]

Morello-Frosch R, Pastor M, Sadd J. 2001 Environmental justice and southern california’s “riskscape”: The distribution of air toxics exposures and health risks among diverse communities. Urban Affairs Review 36:551–578.

Morello-Frosch R, Shenassa ED. 2006 The environmental “riskscape” and social inequality: Implications for explaining maternal and child health disparities. Environ Health Perspect 114:1150–1153. [PubMed: 16882517]

Morello-Frosch R, Jesdale BM, Sadd JL, Pastor M. 2010 Ambient air pollution exposure and full-term birth weight in california. Environ Health 9:44. [PubMed: 20667084]

Myers SL, Lobdell DT, Liu Z, Xia Y, Ren H, Li Y, et al. 2010 Maternal drinking water arsenic exposure and perinatal outcomes in inner mongolia, china. J Epidemiol Community Health 64:325–329. [PubMed: 19692727]

Ochoa-Acuna H, Frankenberger J, Hahn L, Carbajo C. 2009 Drinking-water herbicide exposure in indiana and prevalence of small-for-gestational-age and preterm delivery. Environ Health Perspect 117:1619–1624. [PubMed: 20019915]

Padula AM, Mortimer KM, Tager IB, Hammond SK, Lurmann FW, Yang W, et al. 2014a Traffic-related air pollution and risk of preterm birth in the san joaquin valley of california. Ann Epidemiol 24:888–895e884.

Padula AM, Noth EM, Hammond SK, Lurmann FW, Yang W, Tager IB, et al. 2014b Exposure to airborne polycyclic aromatic hydrocarbons during pregnancy and risk of preterm birth. Environ Res 135:221–226. [PubMed: 25282280]

Rappazzo KM, Daniels JL, Messer LC, Poole C, Lobdell DT. 2015a Exposure to elemental carbon, organic carbon, nitrate, and sulfate fractions of fine particulate matter and risk of preterm birth in new jersey, ohio, and pennsylvania (2000–2005). Environ Health Perspect 123:1059–1065. [PubMed: 25910280]

Rappazzo KM, Messer LC, Jagai JS, Gray CL, Grabich SC, Lobdell DT. 2015b The associations between environmental quality and preterm birth in the united states, 20002005: A cross-sectional analysis. Environ Health-Glob 14.

Ritz B, Wilhelm M, Hoggatt KJ, Ghosh JK. 2007 Ambient air pollution and preterm birth in the environment and pregnancy outcomes study at the university of california, los angeles. Am J Epidemiol 166:1045–1052. [PubMed: 17675655]

Rodrigues T, Barros H. 2008 Maternal unemployment: An indicator of spontaneous preterm delivery risk. Eur J Epidemiol 23:689–693. [PubMed: 18716884]

Romero R, Dey SK, Fisher SJ. 2014 Preterm labor: One syndrome, many causes. Science 345:760–765. [PubMed: 25124429]

Sadd JL, Pastor M, Morello-Frosch R, Scoggins J, Jesdale B. 2011 Playing it safe: Assessing cumulative impact and social vulnerability through an environmental justice screening method in the south coast air basin, california. International journal of environmental research and public health 8:1441–1459. [PubMed: 21655129]

Sagiv SK, Mendola P, Loomis D, Herring AH, Neas LM, Savitz DA, et al. 2005 A time-series analysis of air pollution and preterm birth in pennsylvania, 1997–2001. Environ Health Persp 113:602–606.

Schaaf JM, Liem SMS, Mol BWJ, Abu-Hanna A, Ravelli ACJ. 2013 Ethnic and racial disparities in the risk of preterm birth: A systematic review and meta-analysis. Am J Perinat 30:433–450.

Schifano P, Lallo A, Asta F, De Sario M, Davoli M, Michelozzi P. 2013 Effect of ambient temperature and air pollutants on the risk of preterm birth, rome 2001–2010. Environ Int 61:77–87. [PubMed: 24103349]

Stagnaro-Green A 2009 Maternal thyroid disease and preterm delivery. J Clin Endocrinol Metab 94:21–25. [PubMed: 18984665]

Stayner LT, Almberg K, Jones R, Graber J, Pedersen M, Turyk M. 2017 Atrazine and nitrate in drinking water and the risk of preterm delivery and low birth weight in four midwestern states. Environ Res 152:294–303. [PubMed: 27816866]

Sun XL, Luo XP, Zhao CM, Ng RWC, Lim CED, Zhang B, et al. 2015 The association between fine particulate matter exposure during pregnancy and preterm birth: A meta-analysis. Bmc Pregnancy Childb 15.

Suzuki R, Shimodaira H. 2006 Pvclust: An r package for assessing the uncertainty in hierarchical clustering. Bioinformatics 22:1540–1542. [PubMed: 16595560]

Vigeh M, Yokoyama K, Seyedaghamiri Z, Shinohara A, Matsukawa T, Chiba M, et al. 2011 Blood lead at currently acceptable levels may cause preterm labour. Occup Environ Med 68:231–234. [PubMed: 20798002]

Ward MH, Kilfoy BA, Weyer PJ, Anderson KE, Folsom AR, Cerhan JR. 2010 Nitrate intake and the risk of thyroid cancer and thyroid disease. Epidemiology 21:389–395. [PubMed: 20335813]

Wei T, Simko V. 2017 R package "corrplot": Visualization of a correlation matrix (version 0.84)

Wheeler S, Maxson P, Truong T, Swamy G. 2018 Psychosocial stress and preterm birth: The impact of parity and race. Matern Child Health J

Wilhelm M, Ghosh JK, Su J, Cockburn M, Jerrett M, Ritz B. 2011 Traffic-related air toxics and preterm birth: A population-based case-control study in los angeles county, california. Environ Health 10:89. [PubMed: 21981989]

Zeka A, Melly SJ, Schwartz J. 2008 The effects of socioeconomic status and indices of physical environment on reduced birth weight and preterm births in eastern massachusetts. Environ Health 7:60. [PubMed: 19032747]

Author Manuscript
Author Manuscript
Author Manuscript

Highlights

- Arsenic and Nitrate contamination in drinking water may increase preterm birth risk
- Spatial variability in preterm birth rate may be linked to county urbanization level
- Small exposure signals in a large cohort may involve sizeable population impacts
- Connecting CalEnviroScreen to other database can be instrumental in exposure research

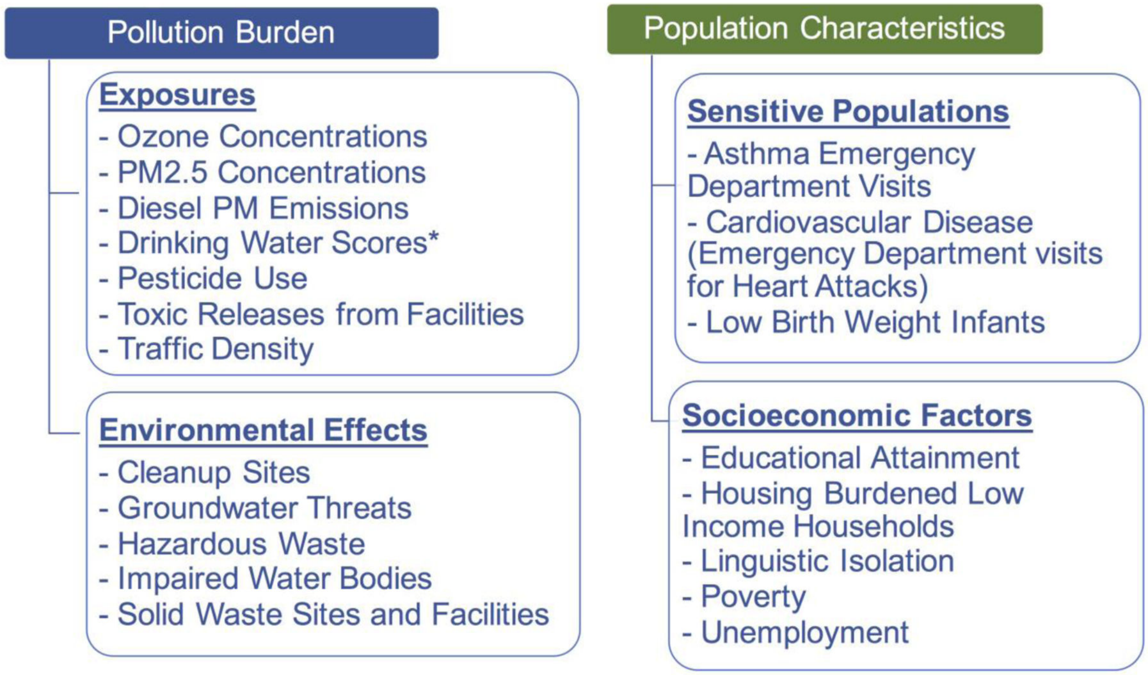

Figure 1.

Categories of Variables in CalEnviroScreen 3.0

Note: This figure was adapted from CalEnviroScreen 3.0 Report (Faust et al. 2017). Further information on the construction of the individual metrics is given in the report.

**Drinking Water Score* contain information of individual contaminants in drinking water (more details can be found in Supplementary Material Table S-1).

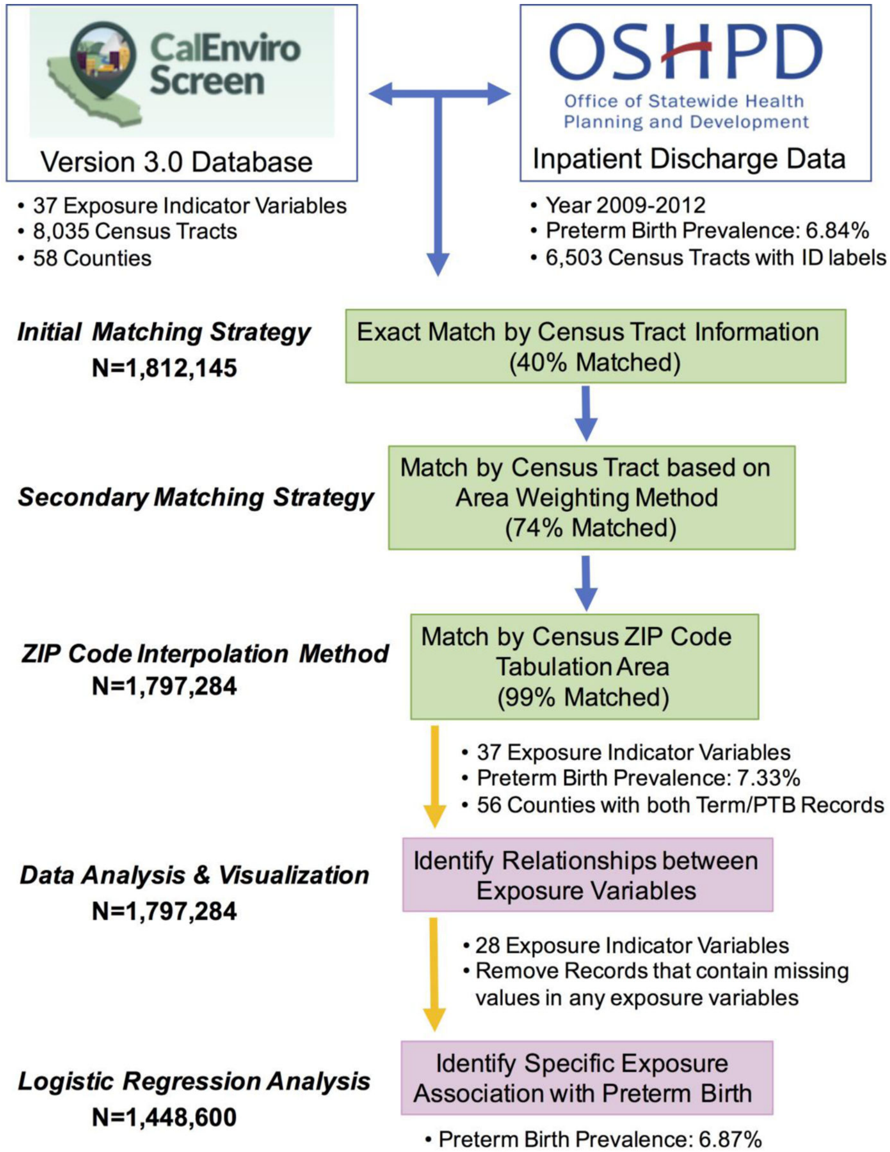

Figure 2.

Flow Chart of Construction of Study Population that Includes Data Matching of CalEnviroScreen and Office of Statewide Health Planning and Development Inpatient Discharge Data

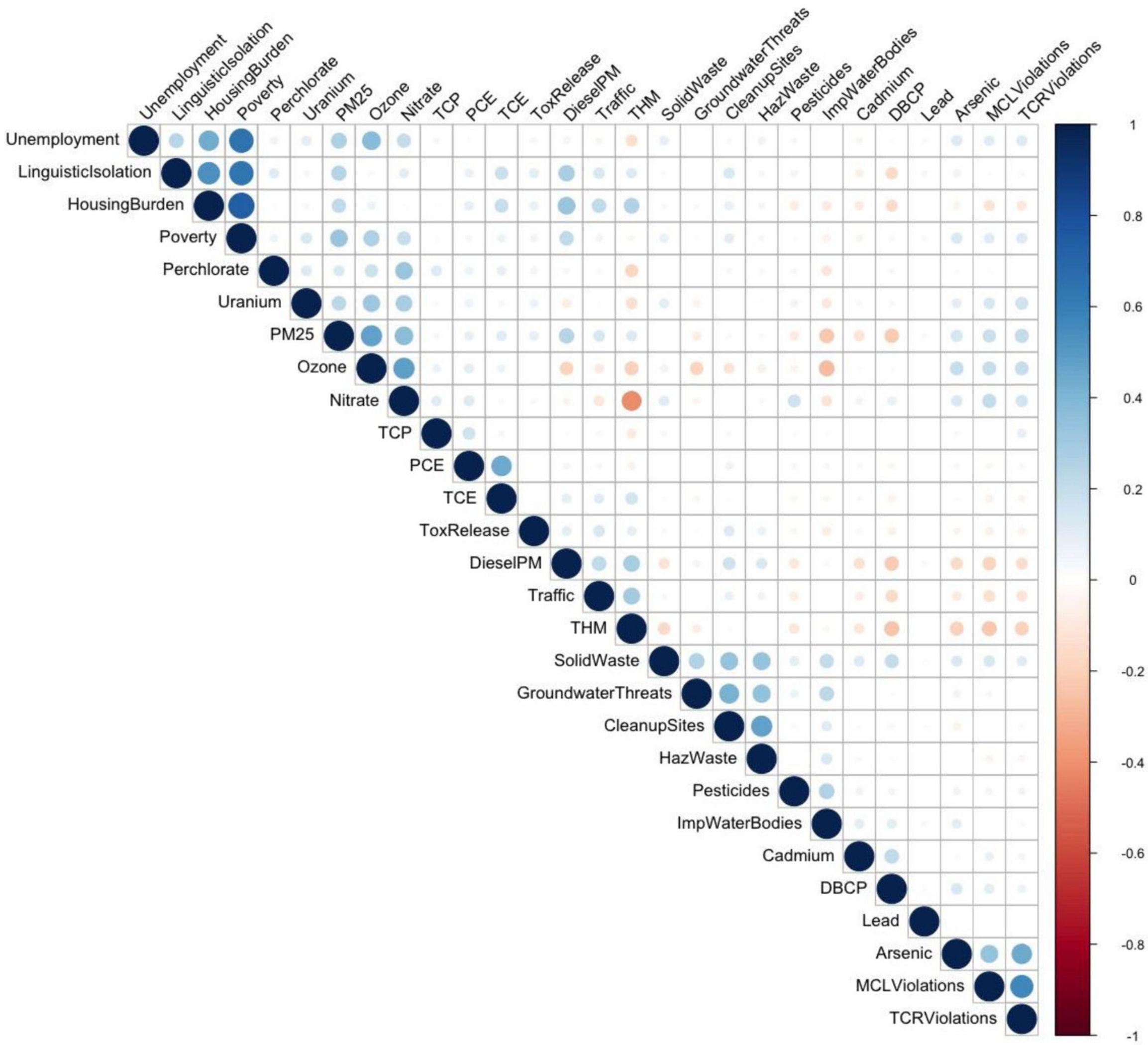

Figure 3.

Correlation Plot of Contaminant Exposure Levels in the Matched Data (N=1,797,284)
DBCP - 1,2-Dibromo-3-chloropropane; MCL- Maximum Contaminant Level; PCE - Tetrachloroethylene; TCE - Trichloroethylene; TCP - 1,2,3-trichloropropane; TCR - total coliform rule; THM – trihalomethane.

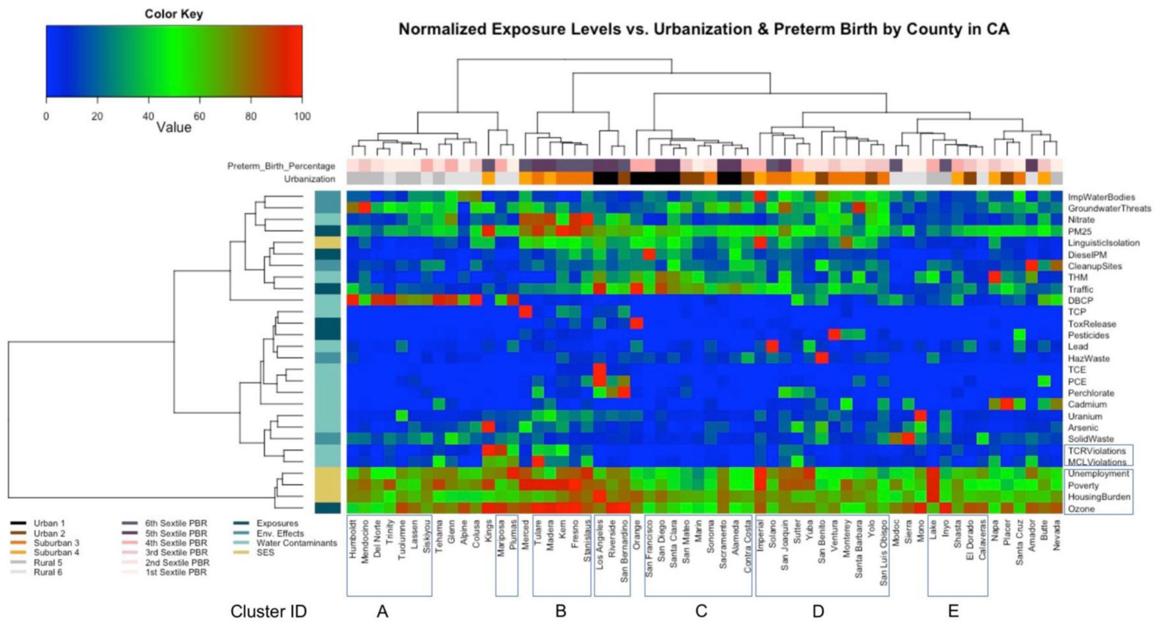

Figure 4.

Heatmap of County-Averaged Contaminant Exposure Levels in California (N=1,797,284)

This heatmap demonstrates clusters of county-wise average estimates of various stressor variables evaluated in this study. Each column represents a county and each row a variable. Warmer color means higher values, or more pollution. The two sidebars at the top of the heatmap indicate the preterm birth percentage and county urbanization respectively. Deeper colors suggest higher preterm birth rate and higher degree of urbanization. The left sidebar shows which category each stressor variable belongs to. Environmental exposure, effects and drinking water contaminants are colored in Indigo, Turquoise and Shallows, but socioeconomic variable in golden sand. The clusters highlighted by blue rectangular box are statistically significant (p-value ≤ 0.05).

PBR – Preterm Birth Rate; SES – Socioeconomic Status. Urban-rural classification was based on 2006 NCHS Urban- Rural Classification Scheme for Counties codes for every county and county equivalent entity in the United States.

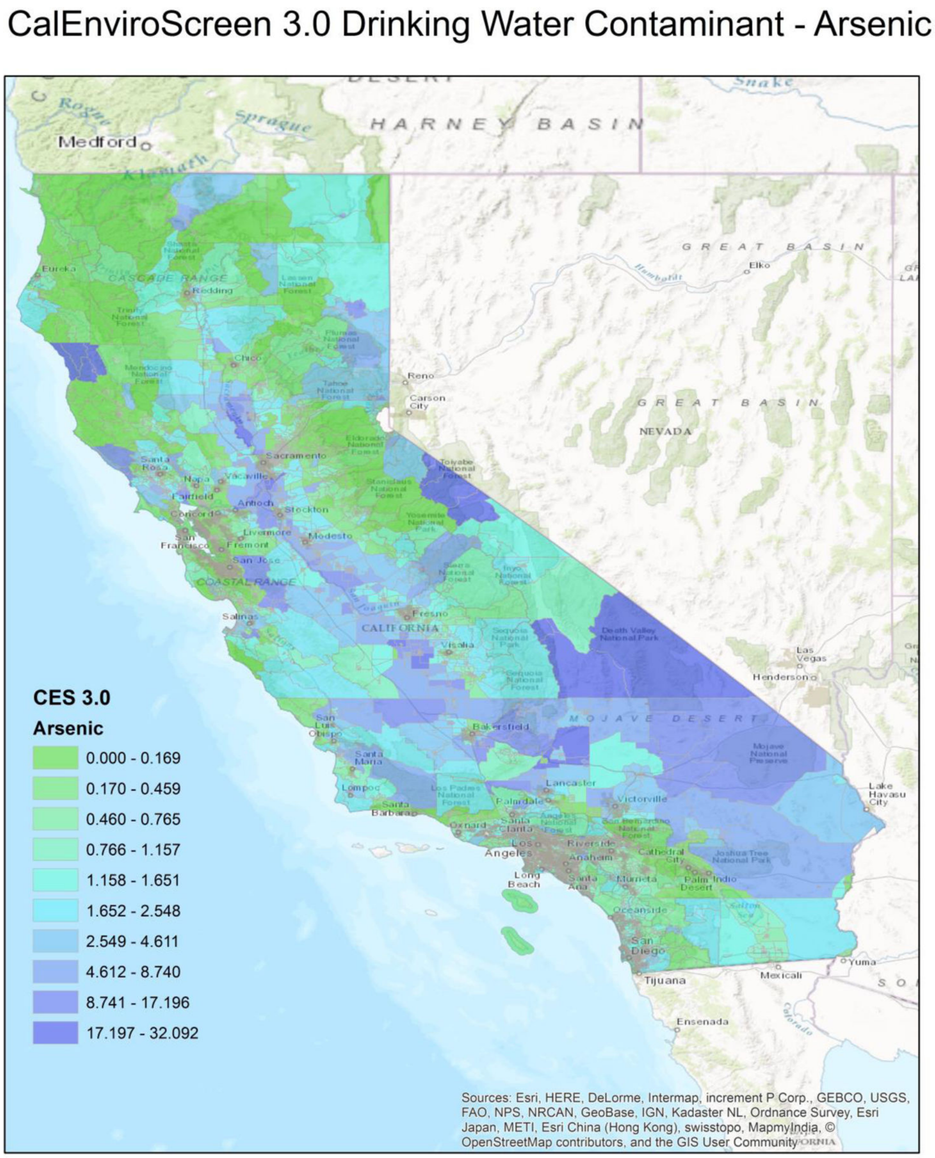

Figure 5.

Spatial Distribution of Arsenic Levels (ppb) in Drinking Water in California Year 2005–2013

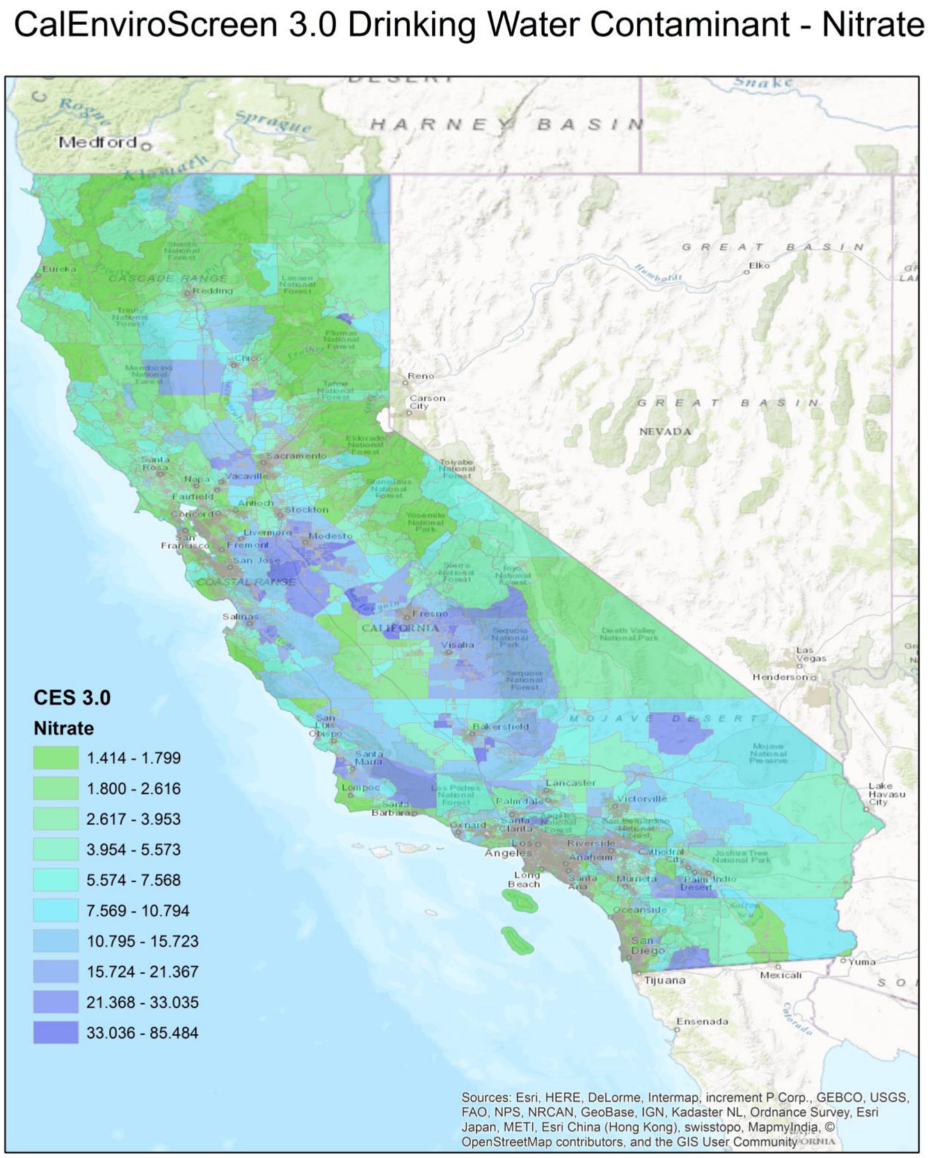

Figure 6.

Spatial Distribution of Nitrate Levels (ppm) in Drinking Water in California Year 2005–2013

Table 1.

Population characteristics in California, 2009–2012 (N= 1,448,600)

Characteristics	n (%)
Race/ethnicity	
White non-Hispanic	367,317 (25.36)
Hispanic	707,336 (48.83)
Black	83,658 (5.78)
Asian	196,503 (13.57)
Other Race	93,786 (6.47)
Maternal age at delivery (years)	
< 18	38,941 (2.69)
18 to 34	1,146,026 (79.11)
> 34	263,633 (18.20)
Maternal education (years)	
< 12	341,931 (23.60)
12	379,457 (26.19)
> 12	727,212 (50.20)
Payer for delivery costs	
Private insurance	663,994 (45.84)
Other payer	784,606 (54.16)

Table 2.

Estimate Interquartile Odds Ratios for Individual Contaminant Exposure Variables associated with PTB (Adjusted for maternal race/ethnicity, maternal age, maternal education and payment of delivery costs).

	p value	Range	IQR exposure	OR for IQR	95% C.I.
$PM_{2.5}$	<0.0001	0 – 19.6	3.35	1.0268	1.0148 – 1.0389
Lead	<0.0001	0 – 1506.34	0.06	1.0001	1.0001 – 1.0002
THM	<0.0001	0 – 96.8	36.77	0.9759	0.9602 – 0.9917
Linguistic Isolation	<0.0001	0 – 67.9	12.60	0.9572	0.9450 – 0.9696
Poverty	<0.0001	0 – 95.2	30.70	1.0976	1.0719 – 1.1239
Unemployment	<0.0001	0 – 76.02	5.96	1.0228	1.0110 – 1.0348
Toxic Release	0.01	0 – 842751.33	4028.72	0.9973	0.9952 – 0.9994
Impaired Water Bodies	0.01	0 – 33	5.92	0.9869	0.9770 – 0.9968
Arsenic	0.01	0 – 32.09	1.38	1.0072	1.0017 – 1.0127
Nitrate	0.01	0 – 85.48	9.33	1.0166	1.0046 – 1.0287
Solid Waste	0.10	0 – 71.5	4.39	1.0062	0.9987 – 1.0137
DBCP	0.13	0 – 1.29	0.01	0.9995	0.9989 – 1.0001
Perchlorate	0.17	0 – 5.36	0.01	0.9999	0.9997 – 1.0000
Housing Burden	0.18	0 – 58.09	10.90	1.0101	0.9949 – 1.0256
Cleanup Sites	0.36	0 – 323.75	11.26	1.0025	0.9970 – 1.0080
Ozone	0.44	0 – 0.07	0.02	0.9945	0.9803 – 1.0089
PCE	0.52	0 – 16.43	0.09	0.9997	0.9990 – 1.0005
Traffic	0.60	0 – 45687.87	696.63	0.9987	0.9937 – 1.0037
Diesel PM	0.65	0 – 253.73	15.81	1.0018	0.9939 – 1.0098
TCE	0.75	0 – 21.49	0.07	0.9999	0.9996 – 1.0003
Uranium	0.82	0 – 137.65	2.15	1.0005	0.9965 – 1.0044
Groundwater Threats	0.91	0 – 1610.09	18.75	0.9997	0.9950 – 1.0045
Hazardous Waste	0.99	0 – 23.36	0.33	1.0000	0.9982 – 1.0018

Note: DBCP – 1,2-Dibromo-3-chloropropane; MCL- Maximum Contaminant Level; PCE – Tetrachloroethylene; TCE – Trichloroethylene; TCP – 1,2,3-trichloropropane; TCR – total coliform rule; THM – trihalomethane. Highlighted rows indicate statistically significant pollutant variables with association directions we expected. The distributions of the variables Pesticide, TCP, TCR Violations, Cadmium and MCL Violations are not amenable to IQR OR calculations, so the corresponding results are not shown.