



## Practice of Epidemiology

# Exposure to Traffic-related Air Pollution During Pregnancy and Term Low Birth Weight: Estimation of Causal Associations in a Semiparametric Model

Amy M. Padula\*, Kathleen Mortimer, Alan Hubbard, Frederick Lurmann, Michael Jerrett, and Ira B. Tager

\* Correspondence to Dr. Amy M. Padula, Department of Epidemiology, School of Public Health, University of California, Berkeley, 1918 University Avenue, Suite 3C, Berkeley, CA 94704 (e-mail: padula@berkeley.edu).

Initially submitted October 27, 2011; accepted for publication February 22, 2012.

Traffic-related air pollution is recognized as an important contributor to health problems. Epidemiologic analyses suggest that prenatal exposure to traffic-related air pollutants may be associated with adverse birth outcomes; however, there is insufficient evidence to conclude that the relation is causal. The Study of Air Pollution, Genetics and Early Life Events comprises all births to women living in 4 counties in California's San Joaquin Valley during the years 2000–2006. The probability of low birth weight among full-term infants in the population was estimated using machine learning and targeted maximum likelihood estimation for each quartile of traffic exposure during pregnancy. If everyone lived near high-volume freeways (approximated as the fourth quartile of traffic density), the estimated probability of term low birth weight would be 2.27% (95% confidence interval: 2.16, 2.38) as compared with 2.02% (95% confidence interval: 1.90, 2.12) if everyone lived near smaller local roads (first quartile of traffic density). Assessment of potentially causal associations, in the absence of arbitrary model assumptions applied to the data, should result in relatively unbiased estimates. The current results support findings from previous studies that prenatal exposure to traffic-related air pollution may adversely affect birth weight among full-term infants.

air pollution; confounding factors (epidemiology); infant, low birth weight; pregnancy

Abbreviations: CI, confidence interval; D/S/A, deletion/substitution/addition; LBW, low birth weight; SAGE, Study of Air Pollution, Genetics and Early Life Events; SES, socioeconomic status; T-MLE, targeted maximum likelihood estimation.

Ambient air pollution is recognized as an important health problem in the United States and around the world (1). Motor vehicles are a major source of ambient air pollution in the United States. Although progress has been made in reducing emissions from individual vehicles, the numbers of vehicles and miles traveled in the United States have grown substantially in the past 15 years (2). Expansion of metropolitan areas (urban sprawl) has increased the travel distances between residential and commercial sites, and the automobile is the primary means of travel. This growth and change in land use has increased the relative contribution of traffic to the urban pollution mixture. Epidemiologic analyses suggest that prenatal exposure to traffic-related air pollutants may be associated with a

variety of health effects, including adverse birth outcomes. Traffic-related air pollution has been associated with intra-uterine mortality (3), low birth weight (LBW) (4, 5), preterm birth (6), small size for gestational age (7, 8), neonatal mortality (9), and postnatal mortality (10, 11). However, there is insufficient evidence to conclude that the relation between traffic exposure and birth outcomes is causal (2). Three general issues may be partly responsible for some of the inconsistency in the findings: 1) definition of the birth outcome, 2) exposure assignment, and 3) statistical methods.

Many studies have investigated multiple birth outcomes without clear specification of a hypothesis related to a single adverse birth outcome. LBW is classified as birth

weight less than 2,500 g. In 2006, 8.3% of infants in the United States were born LBW (12), as compared with 6.7% in 1984 (13). The highest prevalence of LBW is reported for African Americans (11.9%) (12). Most LBW infants are preterm (i.e., born before 37 weeks' gestation). "Term LBW" pertains to infants who weigh less than 2,500 g and are born at or after 37 weeks' gestation, as determined from the last menstrual period, and accounts for approximately 2% of all births. These LBW births often occur because of impaired fetal growth rather than prematurity, thereby suggesting that the causal mechanisms are different.

Previous studies on maternal smoking and LBW suggest that air pollution exposure could potentially affect fetal growth and provide a framework for a potential mechanism of action (14, 15). Therefore, in this study, we evaluated term LBW as a marker of fetal growth separate from the influence of length of gestation. Restriction to full-term infants, in this case, allowed for investigation of a more specific research question.

In most epidemiologic analyses, estimated associations from previous studies have been derived from traditional *parametric* regression methods, which are unable to return the most relevant parameter estimates for understanding the public health impact and, more importantly, are based on arbitrary modeling assumptions and thus may result in biased estimates. We undertook the present study to address 3 limitations of earlier studies with a specific hypothesized birth outcome, high-quality exposure information, and statistical methods that a priori estimate marginal measures of association. This results in parameter estimates with greater public health interpretation, while mitigating the bias from model misspecification as much as possible. This paper provides an illustration of how to use targeted maximum likelihood estimation (T-MLE), and it does so data-adaptively using flexible machine learning algorithms.

## MATERIALS AND METHODS

### Study population

The Study of Air Pollution, Genetics and Early Life Events (SAGE) was designed to investigate the influence of exposure to traffic-related air pollution on pregnancy and birth outcomes. Birth certificates from all 2000–2006 births to women living in the 4 most populated counties in the San Joaquin Valley of California (Fresno, Kern, Stanislaus, and San Joaquin) were obtained from the California Department of Health Services (Sacramento, California).

This analysis was limited to full-term singleton births, for numerous reasons. We defined LBW as birth weight less than 2,500 g and term birth as birth at  $\geq 37$  weeks' gestation. There were a number of exclusions used to isolate term LBW as the outcome and to exclude other adverse birth outcomes. Infants with gestational lengths greater than 44 weeks were excluded because of concerns about data quality regarding reports of the last menstrual period. Infants with birth weights less than 500 g or more than 5,000 g were also excluded because of the likelihood of complications such as birth defects or maternal diabetes, as

well as data quality issues related to the validity of the weight measurement. Finally, mothers with pregnancy complications such as hypertension, diabetes, or uterine bleeding, as reported on the birth certificate, were excluded based on the assumption that the potential impact of traffic exposure would be far outweighed by the influence of these maternal conditions (4).

The maternal residence locations were geocoded with ArcGIS software (ESRI, Redlands, California). Addresses were corrected with ZP4 software (Semaphore Corporation, Aptos, California) in ArcView and SAS, version 9.2 (SAS Institute Inc., Cary, North Carolina). The exposure metric was an indicator of traffic density, calculated from the distance-decayed annual average daily traffic volumes (16) surrounding the geocoded maternal residences. Roadway link-based traffic volumes were derived from Tele-Atlas/Geographic Data Technology traffic-count data for 2005, using methods similar to those of other health-effects studies (16). Further details about exposure assessment are presented in Web Appendix 1, which is available on the *Journal's* website (<http://aje.oxfordjournals.org/>). We split the traffic density indicator into quartiles to characterize the relation across the exposure distribution. For example, the lowest quartile of traffic density represents locations surrounded by small local roads, and the highest quartile is characterized by locations near freeways.

The variables entered into this analysis include: maternal age (<20, 20–35, or >35 years), maternal race (white, Hispanic, African-American, Asian, or other), maternal education (no high school, some high school, some college, or bachelor's or other degree), parity (0 or  $\geq 1$ ), prenatal care (initiation in the first, second, or third trimester), Medi-Cal payment of birth expenses, infant sex, year of birth (2000–2006), and maternal county of residence (Fresno, Kern, Stanislaus, or San Joaquin). Data on these variables were obtained from the birth certificates.

Low socioeconomic status (SES), such as poverty and unemployment, has been associated with adverse birth outcomes (17). Furthermore, SES has been identified as an effect modifier in the relation between air pollution and adverse birth outcomes (18–21). Based on measures implemented in the study by Ponce et al. (21), we created an indicator variable for low SES that was defined as unemployment greater than 10%, more than 15% of households receiving public assistance, and more than 20% of families living below the federal poverty line at the block-group level in the 2000 US Census (21, 22). This variable may not pertain directly to any individual but is meant to provide contextual information about the neighborhoods in which the SAGE study population lived. This research was approved by the University of California, Berkeley, Office for Protection of Human Subjects and the California State Committee for the Protection of Human Subjects.

### Statistical analysis

T-MLE provides a marginal (population-level) estimate and a parameter of interest with a straightforward interpretation (risk difference). It can also be applied to estimating a parameter akin to a causal attributable risk (see literature

on the population intervention models (17, 23)). The parameter of interest is defined as a data-generating distribution on potential outcomes relative to potential interventions (24); the potential outcomes have been referred to as counterfactuals, and we refer the reader to the literature for a more thorough discussion of counterfactuals (25). Briefly, counterfactuals are the set of possible outcomes that would be observed under each possible treatment if, contrary to fact, each person could be observed after exposure to each level of the treatment (i.e., traffic density).

Our goal was to estimate, at the population level, the predicted probability of term LBW had everyone been exposed to each quartile of traffic density. Further details on the statistical methods used are available in Web Appendix 2. In short (under assumptions), the predicted probability of term LBW can be written as

$$EY_a = E_w\{E(Y|A = a, W)\},$$

where  $Y_a$  is the counterfactual outcome had everyone received the exposure  $A = a$ ,  $Y$  is the outcome in the observed data (LBW),  $A$  is (quartile of) traffic density exposure during pregnancy,  $W$  is the vector of potential confounders, and  $E_w$  is the expectation taken over all covariate patterns—that is, the predicted probability of the outcome weighted by the distribution of covariates. This equality is based on the relevant identifiability assumptions (26).

T-MLE consists of 2 modeling steps. The first step is an initial model of the regression  $Q_n^0(A, W) \equiv \hat{E}(Y|A, W)$ . The second step is an augmentation of the initial fit to (heuristically) account for any residual confounding due to the model selection inherent in the machine learning algorithm used to estimate  $Q_n^0$ , which can be achieved simply in this case by adding a “special covariate.” This covariate is a function of the estimated probability of receiving the exposure of interest conditional on the covariates, or specifically

$$\text{logit}(Q_n^1(A, W)) = \text{logit}(Q_n^0(A, W)) + \varepsilon_n \frac{I(A = a)}{g_n(a|W)},$$

where  $g(a|W) = P(A = a|W)$ ,  $Q_n^0(A, W)$  is the initial fit of the regression,  $I(A = a)$  is the indicator function (1 if  $A = a$ , 0 otherwise),  $\varepsilon_n$  is an estimated coefficient, and the  $n$  subscript means “estimated from the data.” Thus, to obtain the “optimal” model of the mean of the outcome given the exposure of interest and confounders, one 1) obtains the initial fit of the regression, 2) estimates  $g$  via an additional regression,  $g(a|W) = P(A = a|W)$ , and 3) using the initial fit as an offset as shown, regresses the outcome against the covariate,  $I(A = a)/g_n(a|W)$ , by means of logistic regression. Because we wanted to avoid arbitrary modeling assumptions, both the  $Q$  model and the  $g$  model were fitted using the deletion/substitution/addition (D/S/A) algorithm. The D/S/A algorithm combines a flexible and aggressive data-adaptive search with V-fold cross-validation (27) and polynomial basis functions, as well as their tensor products. Finally, we applied T-MLE to estimate a parameter akin to attributable risk (17, 28) to measure the impact of a hypothetical intervention that would change the exposure in the

population. In our case, there is no model for the parameter because we are only interested in one comparison with the population mean. This method can be generalized to other situations where one would want to know the difference in mean changes as the level of the intervention changes. Specifically, we estimated the difference in the mean outcome in a population under 1 treatment-specific counterfactual relative to the mean outcome actually observed in the population:

$$\psi \equiv EY - EY_a \stackrel{\text{assumptions}}{=} EY - E_w\{E(Y|A = a, W)\}.$$

This is simply estimated, given our final model for  $E(Y|A, W)$ , and using the empirical distribution to estimate the distribution of  $W$  as

$$\psi_n = \frac{1}{n} \sum_{i=1}^n [Y_i - Q_n^1(a, W_i)].$$

This method accounts not only for the strength of the association between the exposure and the outcome but also for the distribution of the exposure and covariates which actually exist in the study population ( $E(Y)$ ) and the nonparametric estimation of the parameter of interest ( $E_w(Y|A = a, W)$ ) (17). Given that the T-MLE estimate is derived to be asymptotically linear with the efficient influence curve for a parameter in a semiparametric model, we derive a robust standard error by simply obtaining the sample variance of the plug-in influence curve.

For comparison purposes, we performed a traditional logistic regression analysis with quartiles of exposure to traffic density (with the first quartile as the reference category) and included all of the same covariates.

Analyses were performed using R software, version 2.10.1 (DSA package, version 3.1.1; R Foundation for Statistical Computing, Vienna, Austria) and SAS, version 9.2. The R code can be found in Web Appendix 3.

## RESULTS

Of the 329,362 births that took place in the San Joaquin Valley during 2000–2006, the SAGE study population was restricted to singleton births (8,387 multiple births were excluded). Additionally, birth records with missing data for birth weight or gestational length, infants weighing less than 500 g or more than 5,000 g, and infants with less than 20 weeks or more than 44 weeks of gestation were also excluded (gestational age: missing data or <20 weeks for 34,250 infants, >44 weeks for 382; birth weight: <500 g for 289 infants, >5,000 g for 492). Records with missing data on traffic density at the maternal residence were excluded ( $n = 12,826$ ). Births with recorded maternal adverse conditions during pregnancy, such as hypertension ( $n = 461$ ), diabetes ( $n = 4,639$ ), or uterine bleeding ( $n = 517$ ), were excluded. We further restricted the study population to gestational durations between 37 weeks and 44 weeks to capture full-term LBW rather than preterm LBW (30,133 births with gestational ages of 20–36 weeks were excluded). Further description of the excluded

**Table 1.** Characteristics of the Study Population ( $n = 237,031$ ), Overall and by Birth Weight, Study of Air Pollution, Genetics and Early Life Events, San Joaquin Valley of California, 2000–2006

Covariate	Total Study Population ( $n = 237,031$ )		Low Birth Weight Infants ( $n = 5,123$ ; 2.2%)		Normal Birth Weight Infants ( $n = 231,908$ ; 97.8%)	
	No.	%	No.	%	No.	%
Maternal age, years						
<20	32,270	13.6	968	18.9	31,302	13.5
20–35	179,819	75.9	3,561	69.5	176,258	76.0
>35	24,942	10.5	594	11.6	24,348	10.5
Maternal race/ethnicity						
Asian	17,738	7.5	584	11.4	17,154	7.4
African-American	11,560	4.9	527	10.3	11,033	4.8
Hispanic	132,605	55.9	2,653	51.8	129,952	56.0
White	71,522	30.2	1,261	24.6	70,261	30.3
Other	3,606	1.5	98	1.9	3,508	1.5
Maternal education						
No high school	28,027	11.8	539	10.5	27,488	11.9
Some high school	124,128	52.4	3,021	59.0	121,107	52.2
Some college	49,412	20.8	975	19.0	48,437	20.9
Bachelor's or other degree	30,090	12.7	452	8.8	29,638	12.8
Missing data	5,374	2.3	136	2.7	5,238	2.3
Birth costs paid by Medi-Cal						
Yes	127,564	53.8	3,110	60.7	124,454	53.7
No	109,467	46.2	2,013	39.3	107,454	46.3
Low socioeconomic status <sup>a</sup>						
Yes	41,745	17.6	1,102	21.5	40,643	17.5
No	195,286	82.4	4,021	78.5	191,265	82.5
Parity						
0	83,819	35.4	2,303	45.0	81,516	35.1
≥1	153,212	64.6	2,820	55.0	150,392	64.9
Sex of infant						
Male	120,456	50.8	2,221	43.4	118,235	51.0
Female	116,575	49.2	2,902	56.6	113,673	49.0

Table continues

subjects and comparison with those included are presented in Web Table 1. The final study population ( $n = 237,031$ ) is described in Table 1. The proportion of LBW infants was higher in women who were under 20 years of age, were of African-American or Asian race, were formerly nonparous (i.e., this was their first birth), and had less education and women whose birth costs were paid by Medi-Cal.

As Table 2 shows, mean traffic densities were higher in Fresno and San Joaquin counties than in Kern and Stanislaus counties. The distribution of covariates across quartiles of exposure to traffic density during pregnancy is shown in Table 3. The variables that were selected by the D/S/A algorithm in both the outcome ( $Q$ ) and exposure ( $g$ ) models are listed in Table 4. Four of the variables were predictive of both the exposure and the outcome: maternal age >35 years, African-American race, education, and trimester of

initiation of prenatal care. Additional variables came into each of the models.

Table 5 shows the observed and adjusted results for the association between traffic density and term LBW. The adjusted results are predicted probabilities of term LBW if, contrary to fact, everyone had been exposed to each quartile of traffic density. Women who lived in areas with higher traffic density had higher proportions of term LBW infants than women who lived in areas with less traffic, after control for potential confounders derived from the birth certificates and neighborhood SES. Specifically, if everyone in SAGE lived in an area surrounded by high traffic density on busy roadways (fourth quartile of traffic density) during pregnancy, there would be a 2.27% (95% confidence interval (CI): 2.16, 2.38) probability of term LBW in the population, as compared with the 2.02% (95% CI: 1.90,

Table 1. Continued

Covariate	Total Study Population (n = 237,031)		Low Birth Weight Infants (n = 5,123; 2.2%)		Normal Birth Weight Infants (n = 231,908; 97.8%)	
	No.	%	No.	%	No.	%
Initiation of prenatal care						
First trimester	192,905	81.4	3,887	75.9	189,018	81.5
Second trimester	32,676	13.8	870	17.0	31,806	13.7
Third trimester	7,317	3.1	193	3.8	7,124	3.1
Unknown	4,133	1.7	173	3.4	3,960	1.7
Year of birth						
2000	30,788	13.0	608	11.9	30,180	13.0
2001	31,707	13.4	675	13.2	31,032	13.4
2002	32,534	13.7	688	13.4	31,846	13.7
2003	33,082	14.0	717	14.0	32,365	14.0
2004	34,331	14.5	700	13.7	33,631	14.5
2005	35,567	15.0	816	15.9	34,751	15.0
2006	39,022	16.5	919	17.9	38,103	16.4
County of maternal residence						
Fresno	77,093	32.5	1,757	34.3	75,336	32.5
Kern	56,318	23.8	1,253	24.5	55,065	23.7
San Joaquin	59,680	25.2	1,293	25.2	58,387	25.2
Stanislaus	43,940	18.5	820	16.0	43,120	18.6

<sup>a</sup> Low socioeconomic status was defined as unemployment greater than 10%, more than 15% of households receiving public assistance, and more than 20% of families living below the federal poverty line at the block-group level in the 2000 US Census.

2.12) probability of term LBW had everyone lived on less traveled roads (first quartile). The results showed that the estimated probabilities of LBW were lower in the first and third quartiles and higher in the second and fourth quartiles. The highest quartile of traffic density exposure was associated with significantly higher term LBW in comparison with the lowest quartile; however, the exposure-response relation was not monotonic (Figure 1).

Based on these causal attributable risk estimates, if a population intervention could reduce everyone's traffic density exposure during pregnancy to that of the first quartile, the prevalence of LBW among full-term infants would be 2.02% (95% CI: 1.90, 2.12) rather than 2.16%. The tra-

ditional model provided an odds ratio of 1.08 (95% CI: 1.00, 1.17) when comparing the highest quartile of traffic density with the lowest, after controlling for all of the same covariates. The results according to quartile are presented in Table 6.

## DISCUSSION

In this study, we identified a specific source of exposure (traffic) and a specific outcome (term LBW) over a large geographic area to examine a specific hypothesis. The T-MLE analysis provided a (targeted) semiparametric estimate of the causal association between traffic exposure

Table 2. Traffic Density (Scaled to 2003) in the 4 Largest Counties in the San Joaquin Valley of California, Study of Air Pollution, Genetics and Early Life Events, 2000–2006

County	No. of Births	Proximity-weighted No. of Vehicles per Day			
		Mean (SD)	Median	Interquartile Range <sup>a</sup>	Maximum <sup>b</sup>
All 4 counties	237,031	10,369 (15,539)	4,874	225–13,548	163,810
Fresno	77,093	13,116 (17,607)	6,233	793–18,547	162,983
Kern	56,318	7,385 (11,107)	3,456	41–9,612	144,430
San Joaquin	59,680	11,698 (17,725)	5,937	3–14,921	157,340
Stanislaus	43,940	8,419 (11,787)	3,929	0–10,398	163,810

Abbreviation: SD, standard deviation.

<sup>a</sup> First quartile–third quartile.

<sup>b</sup> The minimum number of vehicles was 0; therefore, the maximum represents the full range.

**Table 3.** Characteristics of the Study Population ( $n=237,031$ ) by Quartile of Traffic Density, Study of Air Pollution, Genetics and Early Life Events, San Joaquin Valley of California, 2000–2006

Covariate	First Quartile ( $n=59,197$ )		Second Quartile ( $n=59,271$ )		Third Quartile ( $n=59,210$ )		Fourth Quartile ( $n=59,353$ )	
	No.	%	No.	%	No.	%	No.	%
Birth weight								
Low (<2,500 g)	1,149	1.9	1,322	2.2	1,232	2.1	1,420	2.4
Normal ( $\geq 2,500$ g)	58,048	98.1	57,949	97.8	57,978	97.9	57,933	97.6
Maternal age, years								
<20	6,878	11.6	7,845	13.2	8,287	14.0	9,260	15.6
20–35	45,014	76.1	44,706	75.5	45,015	76.0	45,084	76.0
>35	7,305	12.3	6,720	11.3	5,908	10.0	5,009	8.4
Maternal race/ethnicity								
Asian	4,582	7.7	4,135	7.0	4,006	6.8	5,012	8.4
African-American	1,826	3.1	2,511	4.2	3,003	5.1	4,220	7.1
Hispanic	32,054	54.2	32,880	55.5	33,892	57.2	33,779	56.9
White	19,631	33.2	18,837	31.8	17,553	29.7	15,501	26.1
Other	1,104	1.9	908	1.5	756	1.3	841	1.4
Maternal education								
No high school	6,915	11.7	6,743	11.4	7,360	12.4	7,009	11.8
Some high school	28,352	47.9	29,826	50.3	31,722	53.6	34,228	57.7
Some college	12,898	21.8	12,670	21.4	11,989	20.3	11,855	20.0
Bachelor's or other degree	9,795	16.6	8,704	14.7	6,663	11.3	4,928	8.3
Missing data	1,237	2.1	1,328	2.2	1,476	2.5	1,333	2.2
Birth costs paid by Medi-Cal								
Yes	27,547	46.5	30,053	50.7	33,024	55.8	36,940	62.2
No	31,650	53.5	29,218	49.3	26,186	44.2	22,413	37.8

Table continues

during pregnancy and term LBW. The results did not show a clear exposure-response relation across the quartiles of traffic density; however, there was a significant difference in the predicted probability of LBW between the highest and lowest quartiles of exposure, showing that higher traffic density is associated with increased probability of LBW. The difference between the T-MLE estimate and the crude observed proportion of term LBW suggests that the analysis adjusted for measured confounding and attenuated the risk difference by a small amount (Table 5). The San Joaquin Valley of California is among the most highly polluted areas in the United States. Future analyses of individual pollutants may provide more information on the relation between background levels of air pollutants and traffic density in the San Joaquin Valley.

Most of the previous studies on traffic-related air pollution and birth outcomes have examined the impact of individual exposure to ambient pollutants, including nitrogen dioxide, particulate matter, and carbon monoxide (15). These investigators have reported associations between individual pollutants and various birth outcomes, but the results have been inconsistent as to which pollutant and which time period are the most critical. Some studies have found associations between various adverse birth outcomes

and proximity to highways as a marker of exposure to traffic (20, 29). In one recent study investigating the relation between traffic density and adverse birth outcomes in Shizuoka, Japan, Kashima et al. (30) found no associations. In an earlier study in Los Angeles County, California, Wilhelm et al. (29) estimated higher odds of term LBW (odds ratio = 1.11, 95% CI: 1.04, 1.18) for persons in the highest quintile of distance-weighted traffic density as compared with those in the lowest quintile, though the exposure-response relation was not consistent. That study is the most comparable to our SAGE study because of its large sample size, its location in California, its use of census block-group SES variables, and the similar methods used for outcome ascertainment and exposure assignment. Some notable differences include the years examined (earlier in Los Angeles, when there was generally more traffic), the location, and the statistical methods.

We restricted the study population to full-term births in order to identify a specific etiologic occurrence; however, term LBW is a rare outcome (approximately 2% of births). Birth weight is an important predictor of infant survival and morbidity (31). LBW is also an important indicator of future health and may play a role in the development of chronic diseases throughout life (32). Existing evidence

Table 3. Continued

Covariate	First Quartile (n = 59,197)		Second Quartile (n = 59,271)		Third Quartile (n = 59,210)		Fourth Quartile (n = 59,353)	
	No.	%	No.	%	No.	%	No.	%
Low socioeconomic status <sup>a</sup>								
Yes	5,654	9.6	9,065	15.3	11,084	19.8	15,942	26.9
No	53,543	90.5	50,206	84.7	48,126	81.3	43,411	73.1
Parity								
0	20,468	34.6	20,910	35.3	20,809	35.1	21,632	36.4
≥1	38,729	65.4	38,361	64.7	38,401	64.9	37,721	63.6
Sex of infant								
Male	30,129	50.9	29,948	50.5	30,059	50.8	30,320	51.1
Female	29,068	49.1	29,323	49.5	29,151	49.2	29,033	48.9
Initiation of prenatal care								
First trimester	48,781	82.4	49,195	83.0	47,456	80.2	47,473	80.0
Second trimester	7,627	12.9	7,428	12.5	8,618	14.6	9,003	15.2
Third trimester	1,823	3.1	1,596	2.7	1,968	3.3	1,930	3.2
Unknown	966	1.6	1,052	1.8	1,168	2.0	947	1.6
Year of birth								
2000	7,194	12.2	7,766	13.1	7,879	13.3	7,949	13.4
2001	7,470	12.6	7,862	13.3	8,127	13.7	8,248	13.9
2002	7,762	13.1	8,208	13.9	8,276	14.0	8,288	14.0
2003	8,183	13.8	8,260	13.9	8,380	14.2	8,259	14.0
2004	8,762	14.8	8,597	14.5	8,489	14.3	8,483	14.3
2005	9,339	15.8	8,951	15.1	8,560	14.5	8,717	14.7
2006	10,487	17.7	9,627	16.2	9,499	16.0	9,409	15.8
County of maternal residence								
Fresno	13,910	23.5	21,155	35.7	16,303	27.5	25,725	43.3
Kern	15,698	26.5	16,488	27.8	14,574	24.6	9,558	16.1
San Joaquin	16,456	27.8	11,023	18.6	15,787	26.7	16,414	27.7
Stanislaus	13,133	22.2	10,605	17.9	12,546	21.2	7,656	12.9

<sup>a</sup> Low socioeconomic status was defined as unemployment greater than 10%, more than 15% of households receiving public assistance, and more than 20% of families living below the federal poverty line at the block-group level in the 2000 US Census.

links impaired prenatal growth with adult illnesses such as non-insulin-dependent diabetes, hypertension, and coronary heart disease (33–35). Gestation constitutes a period of human development in which the fetus is particularly susceptible to toxins contained in air pollution because of the high level of cell proliferation, organ development, and the changing capabilities of fetal metabolism (36, 37). Air pollution may affect maternal respiratory or cardiovascular health and, in turn, impair uteroplacental and umbilical blood flow and transplacental glucose and oxygen transport—all known to be major determinants of fetal growth (36).

There were some potential limitations in this study. There was measurement error in the exposure assignment due to the uncertainties in traffic volumes and geocoded locations and the basic nature of the traffic density metric. There may have been misclassification of exposure in

assigning traffic exposure at the maternal residence to an individual. It is unknown how much time each woman spent at the residence reported on the birth certificate during pregnancy and how that time might have varied during the course of pregnancy and vulnerable periods of fetal development.

Because traffic count data were spatial yet not temporal during the study period, it was not possible to target specific periods of pregnancy; rather, the measure used assumed that the exposure was constant across the entire pregnancy. Therefore, the model did not account for seasonal differences in traffic density or women's activity throughout the year, which may correspond to certain periods of fetal development. For some locations, the density of traffic may have varied by season, which was not accounted for in the 9-month assignments, and seasonal differences in chemical and physical transformations of vehicle exhaust were not

**Table 4.** Covariates Chosen by Means of the Deletion/Substitution/Addition Algorithm for the  $Q^0$  Model (Outcome = Term Low Birth Weight) and the  $g$  Model (Treatment = Quartile of Traffic Density) of the Risk of Low Birth Weight According to Traffic Density, Study of Air Pollution, Genetics and Early Life Events, San Joaquin Valley of California, 2000–2006

Covariate	Outcome Model ( $Q^0$ )	Treatment Model ( $g$ )
Maternal age, years		
<20		
>35	X	X
Maternal race/ethnicity		
Asian	X	
African-American	X	X
Hispanic		X
White		
Other		
County of maternal residence		
Fresno		X
Kern		X
San Joaquin		X
Stanislaus		X
Maternal education	X	X
Birth costs paid by Medi-Cal	X	X
Low socioeconomic status <sup>a</sup>		X
Parity	X	
Sex of infant		
Initiation of prenatal care	X	X
Year of birth		X

<sup>a</sup> Low socioeconomic status was defined as unemployment greater than 10%, more than 15% of households receiving public assistance, and more than 20% of families living below the federal poverty line at the block-group level in the 2000 US Census.

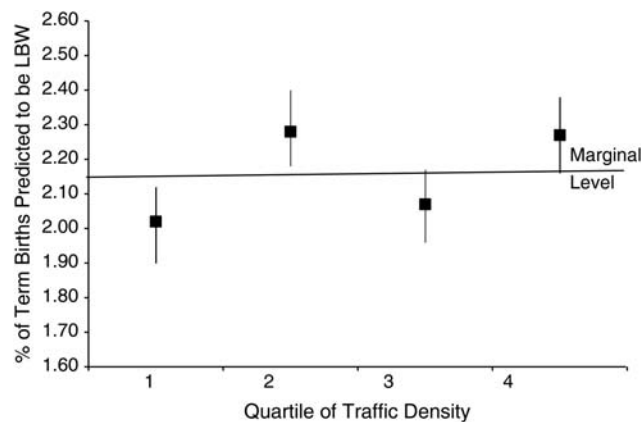
represented in the simple traffic density metric. Although traffic density is an extrapolation of measured traffic activity, it may not represent fully the complex mixture of traffic-

**Table 5.** Observed and Predicted<sup>a</sup> Percentages of Low Birth Weight Among Term Births ( $n = 237,031$ ) by Quartile of Traffic Density, Study of Air Pollution, Genetics and Early Life Events, San Joaquin Valley of California, 2000–2006

Quartile of Traffic Density	Observed %	T-MLE-Estimated %	95% Confidence Interval
1	1.94	2.02	1.90, 2.12
2	2.23	2.28	2.18, 2.40
3	2.09	2.07	1.96, 2.17
4	2.39	2.27	2.16, 2.38

Abbreviation: T-MLE, targeted maximum likelihood estimation.

<sup>a</sup> Obtained using T-MLE analysis.



**Figure 1.** Predicted probability (from targeted maximum likelihood estimation) of term low birth weight (LBW) if all participants were exposed to each quartile of traffic density during pregnancy ( $n = 237,031$ ), Study of Air Pollution, Genetics and Early Life Events, San Joaquin Valley of California, 2000–2006. Bars, 95% confidence interval.

related pollutants contributed by exhaust emissions, brake wear, tire wear, and resuspended road dust.

We were limited to the information that was available on the birth certificate for individual covariates. For example, there was insufficient information on maternal smoking on the birth certificate, which may have changed the results. The prevalence of cigarette smoking among pregnant women in California was 8.7% in 2003 (38), and the association of maternal smoking with birth weight has been well documented (39). If smoking were associated with living near heavily trafficked roadways, it is possible that smoking could confound the relation between traffic-related air pollution and term LBW. Additionally, the data on SES may have been insufficient. We were limited to maternal use of Medi-Cal payment for the birth and block-group-level data from the 2000 census. The data from the birth certificate files that were used in this study were subject to measurement error, and missing information may have biased our results. Although information on birth weight has been found to be reliable in previous studies (92%–100% in comparison with medical records (40)), information on gestational age is less reliable because of variation in maternal recall of the date of the last menstrual period (40). Furthermore, gestational age has a greater proportion

**Table 6.** Odds Ratios<sup>a</sup> for Low Birth Weight Among Term Births ( $n = 237,031$ ) by Quartile of Traffic Density, Study of Air Pollution, Genetics and Early Life Events, San Joaquin Valley of California, 2000–2006

Quartile of Traffic Density	Odds Ratio	95% Confidence Interval
1	1.00	Reference
2	1.10	1.02, 1.19
3	1.01	0.93, 1.09
4	1.08	1.00, 1.17

<sup>a</sup> Obtained using traditional logistic regression.



of missing information, and data may be falsely assumed to be missing at random.

Despite these limitations, the SAGE study population is a large sample with geographic diversity. Most studies of traffic exposure during pregnancy and birth outcomes have included only single metropolitan areas. Although this diversity increases variability and potential confounding, with such a large sample size, this wide geographic area provided us with an opportunity to estimate this relation across a large population with a wider gradient of exposures. Although the association was modest and the outcome is rare, with such a ubiquitous exposure across an entire population, the health impact may be important. Furthermore, this study identified a specific hypothesis regarding the detrimental impact of traffic-related air pollution on fetal growth. If the hypothesis were correct, the implications would go beyond the 2% of infants with term LBW and could apply to all fetuses.

Most previous studies have used only the zip code of the mother's residence and have assigned mothers to a monitoring station within a given distance of the home (4, 41, 42). Our exposure assignment may have been more precise, because the street address was geocoded. Also, unlike most previous studies, in addition to birth certificate characteristics, the SAGE study included neighborhood SES obtained from US census data (19–21).

In this study, we used recently developed semiparametric methods to estimate a causal association. T-MLE provided a marginal (population-level) estimate of the causal association between traffic exposure and term LBW. T-MLE is doubly robust against model misspecification—that is, the estimator will produce unbiased estimates if either the treatment or the outcome mechanism is modeled correctly. Most importantly, T-MLE accounts for heterogeneity in the individual exposure-response relation by targeting the parameter of interest and assumes no particular model for the regression. As in any observational study, reliance on a parametric model means that any estimate has to be interpreted in the context of a misspecified model, that is, with some bias. This approach allows a highly adaptive machine learning algorithm to fit the data (using cross-validation so as to not overfit) and has a plug-in estimator with all of the finite-sample benefits but still augments the model in a manner optimally suited to estimation of the parameter of interest (26).

As with all studies, we made an assumption that oftentimes cannot be tested. For example, we assumed that there was no unmeasured confounding. Our estimates assumed that the correct models were selected, and to maximize the probability that this assumption was valid, we used data-adaptive algorithms for modeling to optimize fit and reduce bias. We tested the experimental treatment assignment assumption by plotting the distribution of the log odds of being exposed to each quartile of traffic density across all quartiles of exposure. The plots revealed that there were no violations of the experimental treatment assignment assumption (i.e., the probability of treatment was not 0 or 1 for any observation); therefore, exposure to specific quartiles of traffic density was not deterministic based on the treatment model. The plots demonstrating this test are

shown in Web Figure 1. The results from the traditional logistic regression analysis showed a similar pattern across quartiles; however, the two analyses were estimating different parameters and are not easily comparable.

In conclusion, the results from these analyses suggest that increased prenatal exposure to traffic may be causally associated with increased risk of term LBW. This study used a measure of a mixture of traffic-related air pollutants and estimated a more causal parameter of interest at the population level. In further studies, researchers should replicate this analysis in other populations, investigate the role of traffic-related air pollution in other adverse birth outcomes such as preterm birth, and gather data on additional covariates.

## ACKNOWLEDGMENTS

Author affiliations: Department of Epidemiology, School of Public Health, University of California, Berkeley, Berkeley, California (Amy M. Padula, Kathleen Mortimer, Ira B. Tager); Department of Biostatistics, School of Public Health, University of California, Berkeley, Berkeley, California (Alan Hubbard); Department of Environmental Health Sciences, School of Public Health, University of California, Berkeley, Berkeley, California (Michael Jerrett); and Sonoma Technology, Inc., Sonoma, California (Frederick Lurmann).

This work was supported by the National Institute for Environmental Health Sciences (grants R21 ESO14891 and P20 ES018173) and the Environmental Protection Agency (grant R834596).

The authors thank Bryan Penfold of Sonoma Technology, Inc., for processing the traffic data and estimating traffic density.

Conflict of interest: none declared.

## REFERENCES

1. World Health Organization. *Tackling the Global Clean Air Challenge*. Geneva, Switzerland: World Health Organization; 2011.
2. Health Effects Institute. *Traffic-related Air Pollution: A Critical Review of the Literature on Emissions, Exposure, and Health Effects*. Boston, MA: Health Effects Institute; 2010.
3. Pereira LA, Loomis D, Conceição GM, et al. Association between air pollution and intrauterine mortality in São Paulo, Brazil. *Environ Health Perspect*. 1998;106(6):325–329.
4. Ritz B, Yu F. The effect of ambient carbon monoxide on low birth weight among children born in Southern California between 1989 and 1993. *Environ Health Perspect*. 1999; 107(1):17–25.
5. Wang X, Ding H, Ryan L, et al. Association between air pollution and low birth weight: a community-based study. *Environ Health Perspect*. 1997;105(5):514–520.
6. Ritz B, Yu F, Chapa G, et al. Effect of air pollution on preterm birth among children born in Southern California between 1989 and 1993. *Epidemiology*. 2000;11(5):502–511.
7. Liu S, Krewski D, Shi Y, et al. Association between gaseous ambient air pollutants and adverse pregnancy outcomes in

- Vancouver, Canada. *Environ Health Perspect.* 2003;111(14):1773–1778.
8. Dejmek J, Solanský I, Benes I, et al. The impact of polycyclic aromatic hydrocarbons and fine particles on pregnancy outcome. *Environ Health Perspect.* 2000;108(12):1159–1164.
  9. Loomis D, Castillejos M, Gold DR, et al. Air pollution and infant mortality in Mexico City. *Epidemiology.* 1999;10(2):118–123.
  10. Woodruff TJ, Grillo J, Schoendorf KC. The relationship between selected causes of postneonatal infant mortality and particulate air pollution in the United States. *Environ Health Perspect.* 1997;105(6):608–612.
  11. Bobak M. Outdoor air pollution, low birth weight, and prematurity. *Environ Health Perspect.* 2000;108(2):173–176.
  12. Martin CR, Brown YF, Ehrenkranz RA, et al. Nutritional practices and growth velocity in the first month of life in extremely premature infants. Extremely Low Gestational Age Newborns Study Investigators. *Pediatrics.* 2009;124(2):649–657.
  13. Hamilton BE, Miniño AM, Martin JA, et al. Annual summary of vital statistics: 2005. *Pediatrics.* 2007;119(2):345–360.
  14. Kannan S, Misra DP, Dvonch JT, et al. Exposures to airborne particulate matter and adverse perinatal outcomes: a biologically plausible mechanistic framework for exploring potential effect modification by nutrition. *Environ Health Perspect.* 2006;114(11):1636–1642.
  15. Shah PS, Balkhair T. Air pollution birth outcomes: a systematic review. Knowledge Synthesis Group on Determinants of Preterm/LBW Births. *Environ Int.* 2011;37(2):498–516.
  16. Kan H, Heiss G, Rose KM, et al. Prospective analysis of traffic exposure as a risk factor for incident coronary heart disease: the Atherosclerosis Risk in Communities (ARIC) study. *Environ Health Perspect.* 2008;116(11):1463–1468.
  17. Hubbard AE, Laan MJ. Population intervention models in causal inference. *Biometrika.* 2008;95(1):35–47.
  18. Yi O, Kim H, Ha E. Does area level socioeconomic status modify the effects of PM<sub>10</sub> on preterm delivery? *Environ Res.* 2010;110(1):55–61.
  19. Zeka A, Melly SJ, Schwartz J. The effects of socioeconomic status and indices of physical environment on reduced birth weight and preterm births in Eastern Massachusetts. *Environ Health.* 2008;7:60. (doi:10.1186/1476-069X-7-60).
  20. Gagnéux M, Auger N, Goneau M, et al. Neighbourhood socioeconomic status, maternal education and adverse birth outcomes among mothers living near highways. *J Epidemiol Community Health.* 2008;62(8):695–700.
  21. Ponce NA, Hoggatt KJ, Wilhelm M, et al. Preterm birth: the interaction of traffic-related air pollution with economic hardship in Los Angeles neighborhoods. *Am J Epidemiol.* 2005;162(2):140–148.
  22. Bureau of the Census, US Department of Commerce. *Census of Population and Housing, 2000. Summary Tape File 3: Technical Documentation.* Washington, DC: US Census Bureau; 2000.
  23. Greenland S. Epidemiologic measures and policy formulation: lessons from potential outcomes. *Emerg Themes Epidemiol.* 2005;2:5. (doi:10.1186/1742-7622-2-5).
  24. Pearl J. *Causality: Models, Reasoning, and Inference.* New York, NY: Cambridge University Press; 2000.
  25. Pearl J. An introduction to causal inference. *Int J Biostat.* 2010;6(2):Article 7. (doi:10.2202/1557-4679.1203).
  26. van der Laan MJ, Rose S. *Targeted Learning: Causal Inference for Observational and Experimental Data.* New York, NY: Springer-Verlag New York; 2011.
  27. Sinisi SE, van der Laan MJ. Deletion/substitution/addition algorithm in learning with applications in genomics. *Stat Appl Genet Mol Biol.* 2004;3:Article 18. (doi:10.2202/1544-6115.1069).
  28. Greenland S, Drescher K. Maximum likelihood estimation of the attributable fraction from logistic models. *Biometrics.* 1993;49(3):865–872.
  29. Wilhelm M, Ritz B. Residential proximity to traffic and adverse birth outcomes in Los Angeles County, California, 1994–1996. *Environ Health Perspect.* 2003;111(2):207–216.
  30. Kashima S, Naruse H, Yorifuji T, et al. Residential proximity to heavy traffic and birth weight in Shizuoka, Japan. *Environ Res.* 2011;111(3):377–387.
  31. Glinianaia SV, Rankin J, Bell R, et al. Particulate air pollution and fetal health: a systematic review of the epidemiologic evidence. *Epidemiology.* 2004;15(1):36–45.
  32. Maisonet M, Correa A, Misra D, et al. A review of the literature on the effects of ambient air pollution on fetal growth. *Environ Res.* 2004;95(1):106–115.
  33. Law C. Fetal influences on adult hypertension. *J Hum Hypertens.* 1995;9(8):649–651.
  34. Barker DJ. Fetal nutrition and cardiovascular disease in later life. *Br Med Bull.* 1997;53(1):96–108.
  35. Barker DJ, Gluckman PD, Godfrey KM, et al. Fetal nutrition and cardiovascular disease in adult life. *Lancet.* 1993;341(8850):938–941.
  36. Ritz B, Wilhelm M. Ambient air pollution and adverse birth outcomes: methodologic issues in an emerging field. *Basic Clin Pharmacol Toxicol.* 2008;102(2):182–190.
  37. Selevan SG, Kimmel CA, Mendola P. Identifying critical windows of exposure for children’s health. *Environ Health Perspect.* 2000;108(suppl 3):451–455.
  38. California Department of Health Services. *Smoking During Pregnancy.* Sacramento, CA: California Department of Health Services; 2005.
  39. Centers for Disease Control and Prevention. *The Health Consequences of Involuntary Exposure to Tobacco Smoke: A Report of the Surgeon General.* Atlanta, GA: Centers for Disease Control and Prevention; 2006.
  40. Northam S, Knapp TR. The reliability and validity of birth certificates. *J Obstet Gynecol Neonatal Nurs.* 2006;35(1):3–12.
  41. Wilhelm M, Ritz B. Local variations in CO and particulate air pollution and adverse birth outcomes in Los Angeles County, California, USA. *Environ Health Perspect.* 2005;113(9):1212–1221.
  42. Ritz B, Wilhelm M, Hoggatt KJ, et al. Ambient air pollution and preterm birth in the environment and pregnancy outcomes study at the University of California, Los Angeles. *Am J Epidemiol.* 2007;166(9):1045–1052.