Association of air pollution and cardiovascular disease rate in california counties

Luqing Ren

Contents

Introduction

Cardiovascular disease (CVD) is a leading cause of death both in California and the United States.The environment problems and source of pollutions have been shown to increase the risk of CVD rate. PM2.5 and other air pollutants such as ozone, nitrogen dioxide, and sulfur dioxide have been considered in the development of health-based standards. Moreover, people living in communities that were identified as "disadvantage" by California Environmental Protection Agency (CalEPA) are more vulnerable to the effects of pollution than others.

The main purpose of this project is to evaluate the association between pollution burden and CVD rate while accounting for community's vulnerability to this association in all California counties.

Methods

Data source

The data set was downloaded from (https://data.ca.gov/dataset/calenviroscreen-3-0-results) through an API. Once downloaded, the desired information was extracted and formed into a data table. This is a data set including environmental, health, and socioeconomic information of all communities in California State. The key independent variables that were examined in this study were **PM2.5**, **ozone**, **traffic density**, **cardiovascular disease rate** and **community category**.

Examine variables

1) **PM2.5** PM2.5 exposure in this data set was an annual mean concentration(ug/m3). After assessing the linearity of PM2.5 and CVD rate, the association was roughly linear.
2) **Ozone** Ozone exposure, which is the mean of summer month (May-October) of the daily maximum 8 hours concentration (ppm). The relationship between ozone exposure and CVD rate was linear.
3) **Traffic density** Traffic exposure(vehicles-km/hr/km) is represented as the number of vehicles (adjusted by road segment lengths in kilometers) per hour per kilometer of roadways.Traffic variable was highly skewed. In order to easily interpret, the traffic density was cut into four levels, and the likelihood test result suggested that keep traffic density variable categorical was better.
4) **The events of cardio_disease** The events of cardio-disease were calculated through multiplying the rate of ED visit per 10,000 residents by the population in that community. Two missing data from population was excluded.

Statistical analysis

An modified Poisson regression model was used to estimate the association of pollutants and cardiovascular disease rate in all communities of California State. Analyses were performed by adjusting for community category confounder. Further model fit assessment was tested by Person chi-square test. The overdispersion of the model was also tested to determine if the variance was larger than what would be expected under a Poisson distribution.Negative binomial regression model was used to address overdispersion. P-values< 0.05 were considered statistically significant.

Results

The relationship between pollution particles and CVD rate

Figure 1: The correlation between cardiovascular disease (CVD) rate and pollution particle: PM2.5 and ozone, by county.

PM2.5 and ozone, are main particles considered in the study. Pollution score is an averaged pollution burden of environmental effects for each county. It is calculated as the average percentiles,the percentile represents a relative score for the indicator, of the environmental effects indicators. The county with higher score therefore has relatively high pollution burdens. In **Figure 1**, PM2.5 and ozone are positively related with CVD rate. The pollution score plot is also showing a positive correlation with CVD rate.

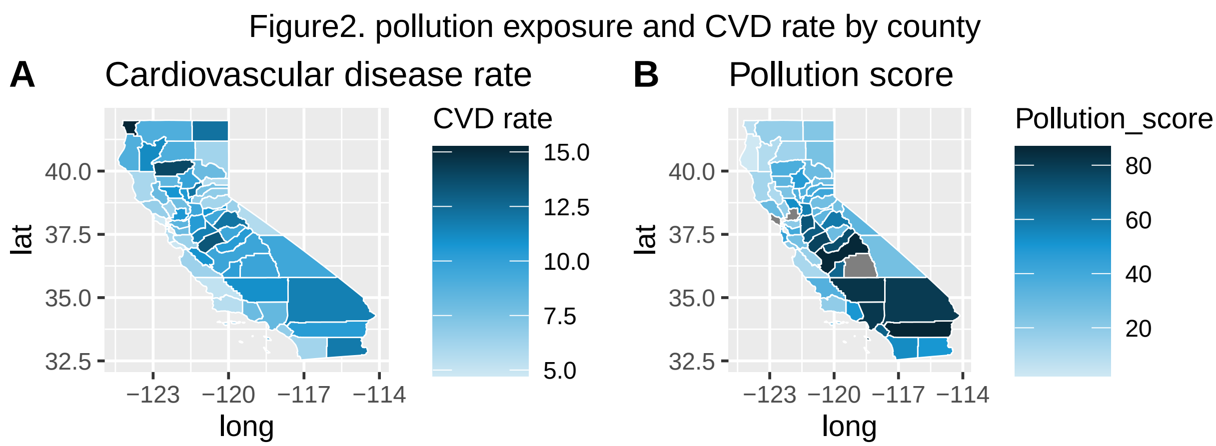

Figure 2: State map depict the pollution score and CVD rate data of each county in California.

Counties with high pollution score is consistent with higher CVD rate. For example, Madera, Fresno, Los Angeles, San Joaquin, San Bernadino and Riverside counties have higher pollution score as well as a high CVD rate(**Figure 2**). However, there are some inconsistency. For instance, the northern counties of state shows lower pollution burden but have relatively higher CVD rate. This suggests that other life characteristics and factors may contribute to their high CVD rate.

Pollution burden and community category

County	Score	Community category
Fresno	7.483	Yes
Stanislaus	7.286	Yes
Tulare	7.246	No
San Bernardino	7.218	Yes
Kern	7.183	Yes
Kings	6.997	Yes
Los Angeles	6.976	Yes
Merced	6.914	Yes
San Joaquin	6.498	Yes
Riverside	6.471	Yes

Table 1: Top 10 counties with high pollution burden and their community category.

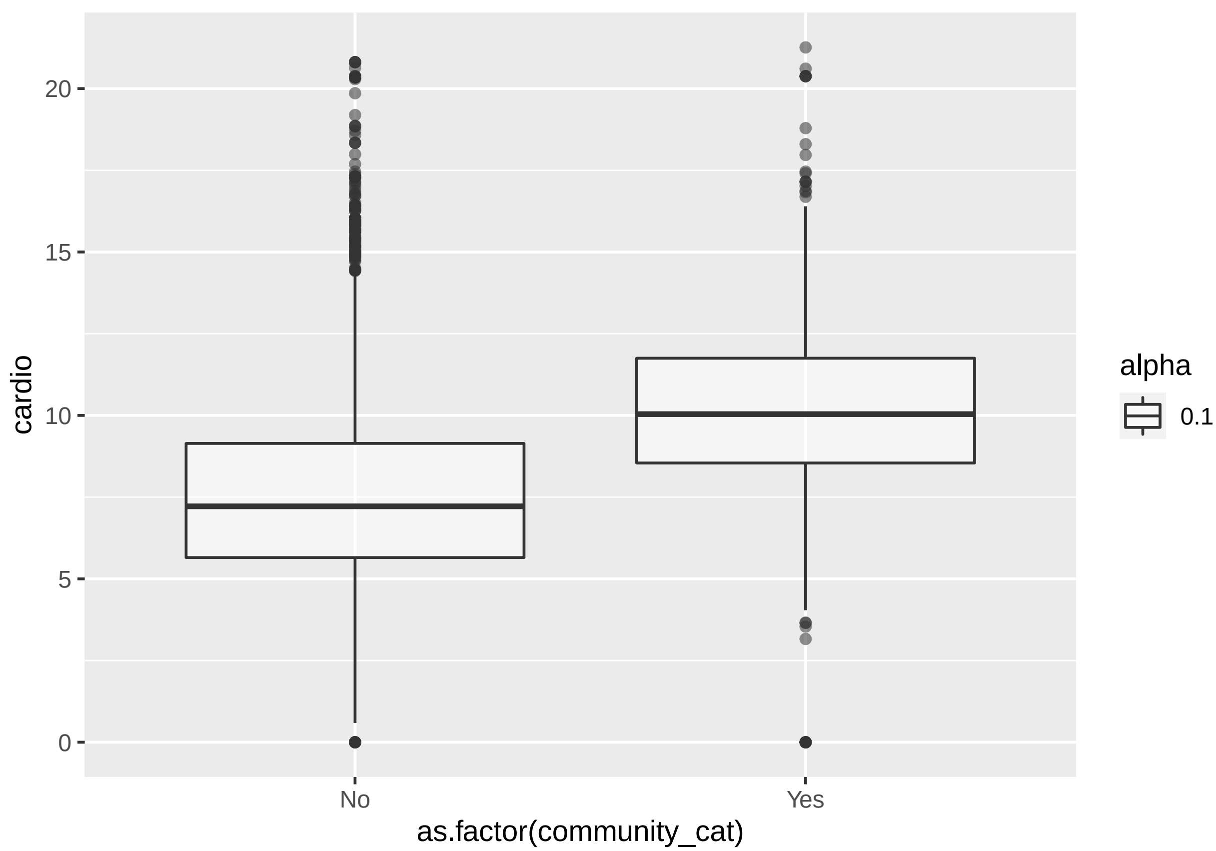

Figure 3 : CVD rate by community category.

Among the top 10 cities with high pollution burden, 90% of them are indisadvantage communities. CalEPA classifies communities as "disadvantage" due to the environmental conditions and vulnerability of people living in those communities. **Figure3** and **table1** both indicate that living in an disadvantageous community may increase the risk of CVD.

CVD rate with pollution burden by community category

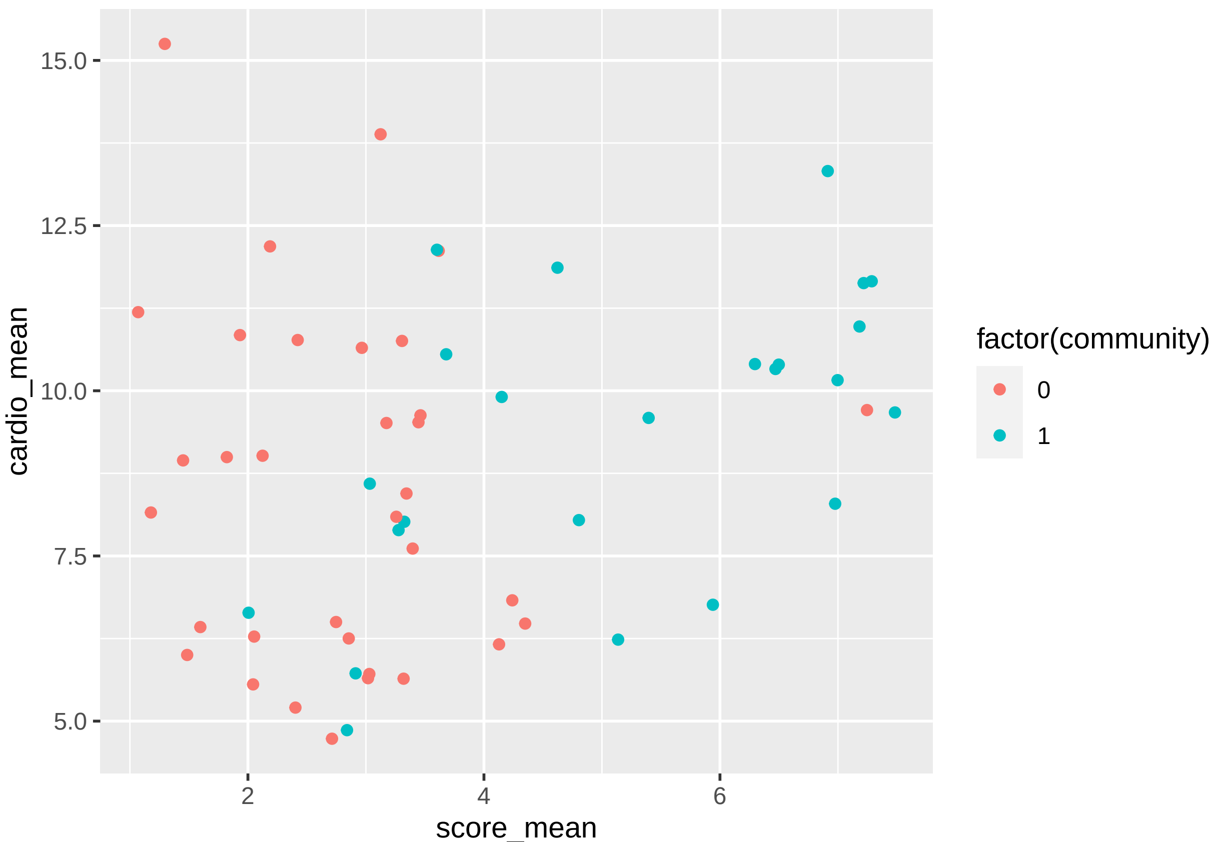

Figure 4 : Cardiovascular rate vs. pollutant burden by community category.

Figure 4 shows the association between CVD rate and pollution burden by community category. The average pollution burden in disadvantage communities is obviously higher than non-disadvantage communities. However, the relationship between CVD rate and pollution burden is not as obvious as in **Figure 1**.

Variable statistic description

variable	n_missing	complete_rate	mean	SD	p0	p25	p50	p75	p100
PM2.5	19	1.00	10.38	2.60	1.65	8.70	10.37	12.05	19.60
ozone	0	1.00	0.05	0.01	0.03	0.04	0.05	0.06	0.07
traffic	56	0.99	943.04	907.36	22.41	442.08	699.89	1190.07	45687.87

Table 2: Characteristics of all variables

The concentration of PM2.5 exposure in this data set varied from 1.65 to 19.6 ug/m3 (**Table 2**). Mean PM2.5 exposure was 10.4 ug/m3. Ozone exposure, varied from 0.03 to 0.07 ppm, and mean ozone exposure was 0.047ppm. The mean traffic exposure (vehicles-km/hr/km) was 943.04 (vehicles-km/hr/km). There are 19 missing data in PM2.5 and 56 missing data in traffic density, which are only 0.2% and 0.7% proportion of whole data set, respectively.

Traffic density level VS PM2.5 and Ozone

Traffic density level (vehicles-km/hr/km)	PM2.5 (ug/m3)	Ozone (ppm)
[22.4,442]	9.975939	0.0489529
(442,700]	10.304066	0.0469590
(700,1.19e+03]	10.572666	0.0471831
(1.19e+03,4.57e+04]	10.618026	0.0465895

Table 3: Traffic density vs PM2.5 and Ozone

The lowest traffic density level was 22.4-442 (vehicles-km/hr/km).The mean PM2.5 and ozone value associated with the lowest traffic density were 9.98(ug/m3) and 0.049(ppm), respectively. The highest traffic density level was associated the highest PM2.5 value 10.62(ug/m3). Thus, PM2.5 was positively associated with traffic density level, while ozone was not.

Association between CVD rate and pollutants

Predictors	RR	2.5 %	97.5 %	p_value
(Intercept)	1.53	1.42	1.66	0.000
PM2.5	1.01	1.01	1.02	0.000
traffic.q4(442,700]	1.07	0.96	1.19	0.214
traffic.q4(700,1.19e+03]	1.16	1.02	1.32	0.025
traffic.q4(1.19e+03,4.57e+04]	1.47	1.24	1.75	0.000
ozone_ctr	2.27	2.15	2.40	0.000
PM2.5:traffic.q4(442,700]	0.99	0.98	1.00	0.173
PM2.5:traffic.q4(700,1.19e+03]	0.98	0.97	1.00	0.008
PM2.5:traffic.q4(1.19e+03,4.57e+04]	0.96	0.94	0.97	0.000

Table 4:PM2.5, ozone, traffic density with the events of cardiovascular disease.

In preliminary model, the CVD rate modeled by PM2.5 exposure was statistically significant(95%CI 1.01,1.02; $p<0.001$), and the CVD rate of PM2.5 was 1.01 times than the base line. Ozone exposure was associated with 2.27 times the rate of CVD(95%CI 2.15,2.41; $p<0.001$). Traffic density was statistically significant with CVD rate, and CVD rate in the highest traffic density level (95%CI 1.23,1.75; $p<0.001$) was 1.47 times the lowest traffic level. PM2.5 exposure has a statistically significant interaction with traffic density($p<0.001$) when traffic density is higher than 700 (vehicles-km/hr/km).

Adjusted association between CVD rate and all variables

Predictors	RR	2.5 %	97.5 %	p_value
(Intercept)	1.74	1.60	1.89	0.000
PM2.5	1.00	0.99	1.00	0.463
traffic.q4(442,700]	1.11	0.99	1.24	0.080
traffic.q4(700,1.19e+03]	1.30	1.13	1.49	0.000
traffic.q4(1.19e+03,4.57e+04]	1.84	1.53	2.21	0.000
ozone_ctr	2.21	2.08	2.35	0.000
community	1.35	1.31	1.39	0.000
PM2.5:traffic.q4(442,700]	0.99	0.98	1.00	0.043
PM2.5:traffic.q4(700,1.19e+03]	0.97	0.96	0.98	0.000
PM2.5:traffic.q4(1.19e+03,4.57e+04]	0.93	0.92	0.95	0.000

Table 5: Adjusted association between the pollution and CVD rate.

After check for overdispersion, there was an 1.14 (P<0.001) overdispersion term in the Poisson model. A negative binomial regression model was performed. In the final negative binomial model,the estimated adjusted baseline CVD rate did not change much after adjust for community category. PM2.5 was not statistically associated with CVD rate after the adjustment(p=0.472), and the exposure had the same CVD rate as non-exposure(95% CI 0.99,1.00;p=0.472). Every one-unit ozone exposure is associated with 2.21 times the CVD rate(95% CI 2.08,2.35;p<0.001), after adjusting for community. The highest traffic density exposure was associated with 1.84 times CVD rate (95% CI 1.53,2.21;p<0.001). The rate of CVD increased 25% compared to unadjustment. The interaction between PM2.5 and traffic density was significant after adjustment.

Conclusion

Pollutants such PM2.5, Ozone and traffic were statistically significantly assocaited with CVD rate and living in an disadvantageous community may increases risk of CVD rate. The baseline of CVD rate after adjustment of community category increased 13%. The CVD rate in highest traffic density level increased 25% compared to unadjustment. Although the CVD rate associated with PM2.5, and ozone exposure did not change much after adjustment, the people with ozone exposure still had 2.21 times the rate of CVD than the baseline. PM2.5 had a significant interaction with traffic density level. Thus, improvements in air quality would be helpfull to reduce over all CVD prevalance acrross california. In additon, efforts to identify the pollution source that accounting for a community's vulnerability would benefitial to those disadvantaged communities which suffered most from CVD occurence.