Appendix A. Matching Zip Codes and Pooling Zip Codes into Zip Code Areas

Because the mortality and morbidity data were available at the zip code level, zip codes were chosen as the geographic unit of analysis. This appendix discusses how zip codes were processed.

A1. Zip Code Sources

Zip code data were available from the Census[1] and the commercial geographic information system (GIS) package ArcGIS. These provided information on which zips were populated and which zips were fully or partly contained the San Francisco Bay Area. A further list from Alameda County Public Health Department[2] (ACPHD) provided data on 1998 zip codes. Although somewhat out of date, the ACPHD list was more comprehensive than the other lists.

The zips from the hospital admissions and emergency room visit data supplied by the Office of Statewide Health Planning and Development (OSHPD) derives from reports by hospitals of patient addresses. The zip codes from mortality data were provided to the State by counties.

A2. Pooling Census and ArcGIS Zips

Zip codes change over time, with altered boundaries, and with new zips being added. Zip code data from Census 2010 and ArcGIS were used to develop a list of zip codes that were populated and totally or partly within the Bay Area. These also provided estimates of the fraction of the zip code within each county. The ArcGIS data included zips labeled "Enclosed Postal", which appear to be zips dedicated to post office boxes. Ignoring these, there was close agreement between the two datasets, with 305 populated zips. The ArcGIS data had information specifying which populated zip each Enclosed Postal zip belonged to. Each enclosed zip was assigned to

[1] The Census Bureau examines the zip codes in each Census block and, if there is more than one zip for that block, assigns the zip corresponding to the greatest population. The resulting set of zip codes are termed Zip Code Tabulation Areas (ZCTAs). Here we've used the ZCTAs from the 2010 census.

[2] Matt Beyers, Community Assessment, Planning, Education, and Evaluation, Alameda County Public Health Department, personal communication.

the populated zip that contained it. The list from ACPHD provided an additional 41 zips, connecting these with the set of populated zips, for a total pooled list of 491.

The OSHPD data had a large number of zip codes, of which all but seven were in in the pooled list. A web search indicated that zips 94096 and 94098 were in San Bruno, zip 94066; zips 94153, 94154, and 94157 were in San Francisco zip 94103; zip 94875 was in Richmond zip 94804; and zip 95171 was in San Jose zip 95111. The mortality data were much simpler and fit completely within the set of zips in the pooled list.

A3. Defining Zip Areas

The pooled list of zip codes linked zips with zero population to an Enclosed Postal zip. The populated zips were sorted by 2010 population and, for zips with less than 2,000 residents; Google Earth was used to merge them with neighboring zips. In some cases, the zip receiving merged neighboring zips had a large population. In other cases, both the merged and receiving zips had populations less than 2,000. In these cases, smaller population zips were merged into larger, until the sum was greater than 2,000. Through this process, we arrived at made total of 244 zip areas.

A3.1 Handling Exceptional Zips

In subsequent analyses of the health of zip code populations, it was clear that several zip areas were exceptional and reasonable to treat as special cases. These were:

1. 94720, the campus of UC Berkeley made up largely of students, average age 21
2. 94535, Travis AF Base made up largely of military personnel, average age 21.
3. 94964, San Quentin, made up largely of prisoners

Of these, San Quentin had 5,094 residents; the others had less than 4,000. Because of their unusual demographics, we folded these zips into neighboring zips: 94720 into 94704 (Berkeley

south of campus, 94535 into 94533 (Vacaville), and 94964 into 94901 (San Rafael).[3] Folding these zips into others, we were left with a final total of 241 zip areas.

[3] These areas are important, but their health dynamics were found to be unique and not easily meshed into the analysis: each of these exceptional zips exhibited very large uncertainties in life expectancy estimates. Another zip—94305, Stanford University—had a young population (average age 26) but does include a substantial number of older people. The University of Santa Clara has its own zip, but this zip had already been merged with the zip surrounding it. Other universities with large student populations have zips that include populations around it. In such situations, the zips were included in the analysis.

Appendix B. Assigning Fine Particulate Matter and Ozone Concentrations to Zip Code Area

B1. Estimating Fine Particulate Matter Concentrations by Zip Code Area

Initial estimates of fine particulate matter ($PM_{2.5}$) concentrations in Bay Area zip codes were derived from regional air quality modeling. The Community Multiscale Air Quality (CMAQ) model was used to simulate direct and secondary $PM_{2.5}$ for two-week periods in January, March, May, August, October, and December, 2010 and 2011 using a 4x4 km grid that included the Bay Area. These $PM_{2.5}$ modeling estimates were adjusted to match observations more closely, as described below. Adjusted concentration estimates were averaged over the period to each Census block in the Bay Area. Each zip code area was then assigned the average of the $PM_{2.5}$ concentrations in the Census blocks contained within it, weighting by Census block population.

To calibrate the CMAQ modeled concentrations, observed $PM_{2.5}$ from 12 Bay Area $PM_{2.5}$ monitoring sites was matched with the $PM_{2.5}$ from the corresponding CMAQ model grid square. The data for each site were limited to the days where both modeled and observed data were present. Table B1 shows the sites and the amount of data available by site. The Point Reyes site is part of the Interagency Monitoring of Protected Visual Environments (IMPROVE) network, with a 1-in-3 day sampling schedule. The other sites are Air District sites, some of which had limited data collection in the months April-September. The Fremont site was closed in December 2010.

Table B1. Measured versus modeled $PM_{2.5}$ at monitoring sites

	Measured			Modeled
Site	n	missing	mean (ug/m3)*	mean (ug/m3)*
Livermore	139	24	9.6	14.3
Oakland	163	0	9.8	15.8
Fremont	64	99	9.1	15.1
Concord	118	45	9.2	13.7
San Rafael	154	9	11.8	9.4
San Francisco	159	4	11.3	22.4
Redwood City	153	10	9.4	14.3
Gilroy	154	9	8.7	15.0
San Jose	115	48	12.1	18.5
Vallejo	123	40	10.3	11.9
Santa Rosa	154	9	8.5	12.1
Pt Reyes	41	122	4.9	5.2

* Means taken over the same set of days for each site, limited to two-week periods in Jan, Mar, May, Aug, Oct, and Dec, 2010 and 2011, and further limited by the number of days with measured $PM_{2.5}$ at each site.

Figure B1 shows measured vs. modeled $PM_{2.5}$ for Livermore, as an example. There is a reasonably good correlation (0.65), at this site but the model clearly tends to overestimate as shown by the majority of points lying below the line y = x. A linear regression line is also shown. Regression lines for all sites are shown in Figure B2. No pattern was found that could be applied over the Bay Area to explain the variation in the regression lines. Instead, we formed a consensus line that was applied to all zip codes, also shown in Figure B2. The equation for the consensus line was

$$\text{adjusted } PM_{2.5} = 0.37 * \text{modeled } PM_{2.5} + 4.$$

This equation was very similar to one obtained from regressing the 12 site measured means vs. the 12 corresponding modeled grid means.

The CMAQ model's tendency to over-predict $PM_{2.5}$ was due, at least in part, to two problems with modeling inputs, which are currently being addressed. First, an overly stable atmosphere in the meteorological input fields generally inhibited mixing. Second, emissions from source categories were overestimated in some areas. California Air Resources Board estimates of off-road vehicle emissions were too high, particularly in San Francisco. There was also likely a problem with an over-estimation of woodsmoke in San Francisco and San Jose. The emissions overestimates led to greater overestimates of $PM_{2.5}$ concentrations in these areas. With these latter considerations in mind, any modeled $PM_{2.5}$ above 18 $\mu g/m^3$ was first reduced to 18 $\mu g/m^3$. The consensus equation was then applied to the mean value in each grid square, then this adjusted mean used as the $PM_{2.5}$ value for a zip code whose centroid was in the grid square.[4]

[4] We also made an exception for Vallejo, where the adjusted modeled value was considerably below the observed. Here we used the actual mean value from the monitoring site.

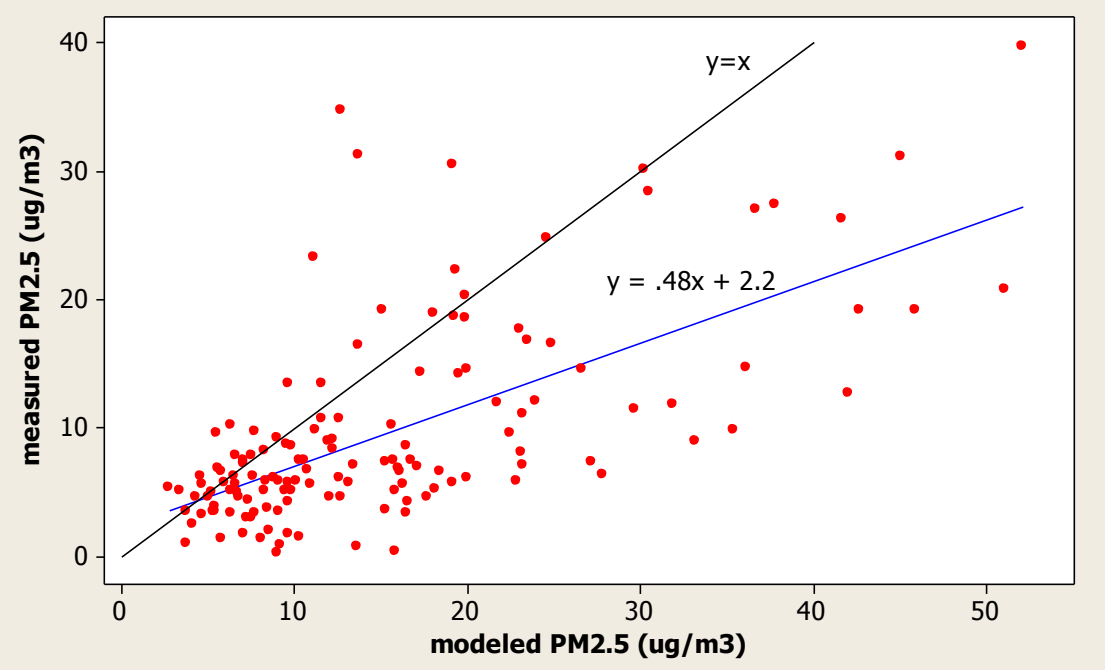


Figure B1. Measured versus modeled $PM_{2.5}$ in Livermore, 2010-2011.

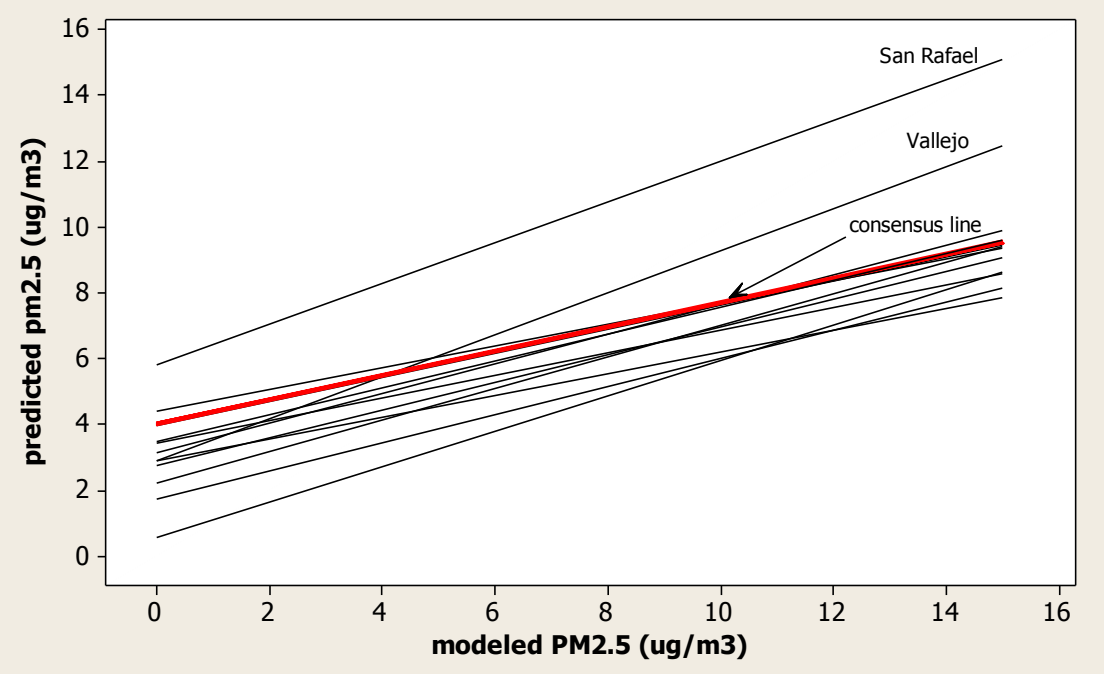


Figure B2. Regression lines: measured versus modeled for 12 Bay Area sites and the consensus line.

B2. Estimating Ozone Concentrations by Zip Code Area

Average annual Ozone concentrations were estimated for each zip code area by interpolation from monitored ozone data.

Choice of ozone indicator

Ozone is collected on an hourly basis, but studies that have investigated ozone-health relationships have generally focused on one of several summary statistics: daily 1-hour maximum, daily 8-hour maximum, or 24-hour average. US EPA's Benefits Mapping and Analysis Program (BenMAP, see Appendix C) has chosen 8-hour maximum ozone as its indicator. It includes studies that used other indicators, but offers an estimated effect estimate (β) to correspond to what the concentration-response (C-R) effect would have been had the study used 8-hour maximum ozone. Thus, for our purposes, 8-hour maximum ozone is used as the variable of interest.

Ozone in the Bay Area is strongly affected by concentrations transported onshore from the Pacific Ocean. There is a constant non-zero background with an hourly median value of around 30 ppb. Here we assume that the Bay Area receives a background 8-hour maximum ozone concentration of 40 ppb, and base our estimate of the impact of anthropogenic ozone on the increments *above* this value. Thus, for example, if the 8-hour ozone in an area were 70 ppb, then a value of 30 ppb (70 – 40) would be used in the C-R function.

Interpolation of 8-hour ozone

Daily maximum 8-hour ozone data were computed for each BAAQMD monitoring site for each day of 2010 and 2011. Because of the powerful influence of offshore ozone on Bay Area concentrations, an additional "pseudo-sites" were included at 4 points offshore from just south of Pescadero to Point Reyes. These sites were assigned the approximate ozone background of 40 ppb.

8-hour ozone concentrations were estimated for each Bay Area Census block as a weighted average of the 8-hour ozone from nearby sites with weights of $1/(d^2 + .5)$, where d = distance between the block group and the site in kilometers. These daily interpolated 8-hour ozone values were then truncated at 40 ppb, and the increments above 40 ppb averaged over the two years.

The zip code area 8-hour ozone was then estimated as the average of the ozone for the blocks within the zip code area weighted by the population of the block. Figure B3 shows that the exposure to elevated ozone is lowest along the coast and in the Oakland-San Francisco area. Highest exposures occur around Livermore and Bethel Island in the east and in the San Martin-Gilroy area in the south.

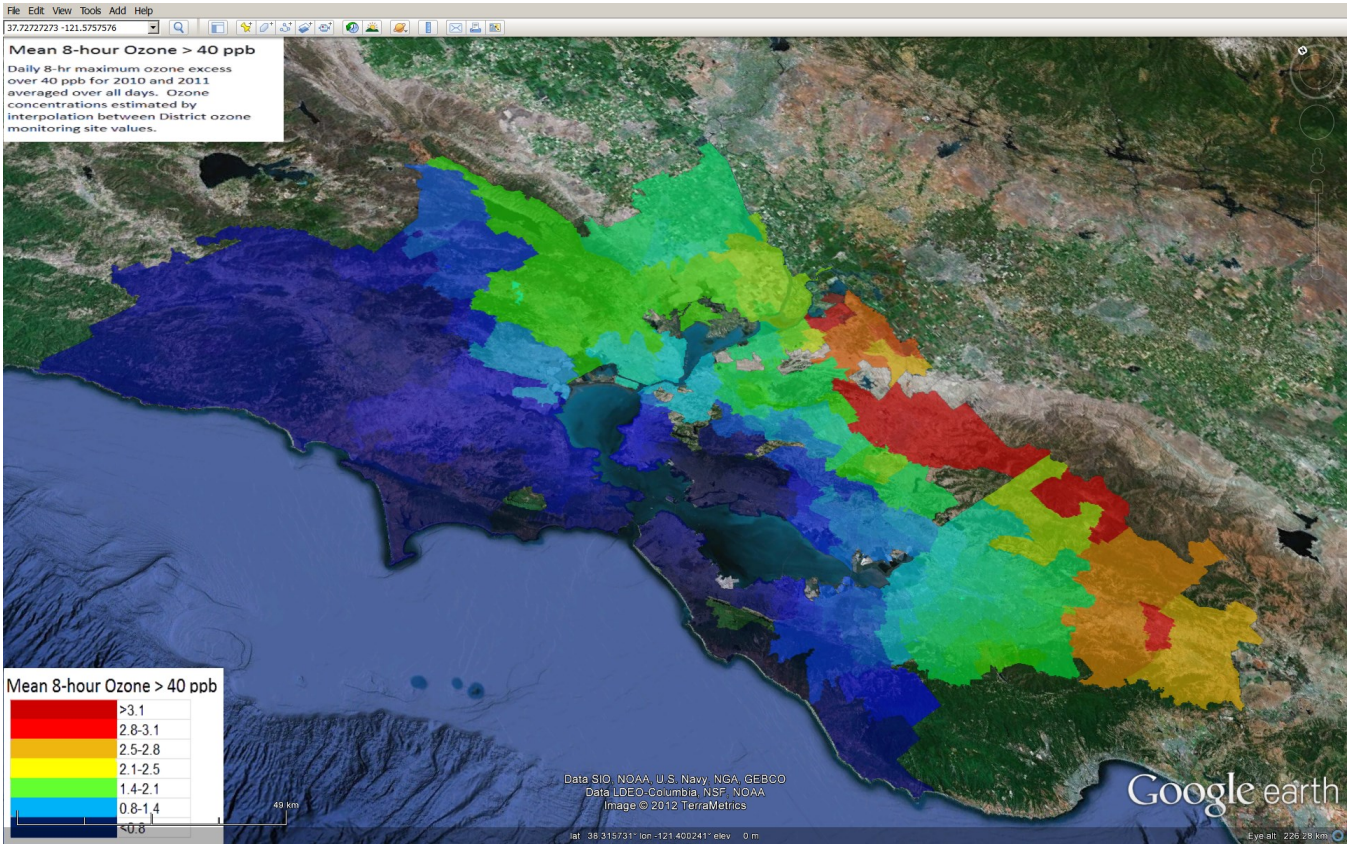


Figure B3. Mean 8-hour ozone (over 40 ppb) mapped to zip code areas.

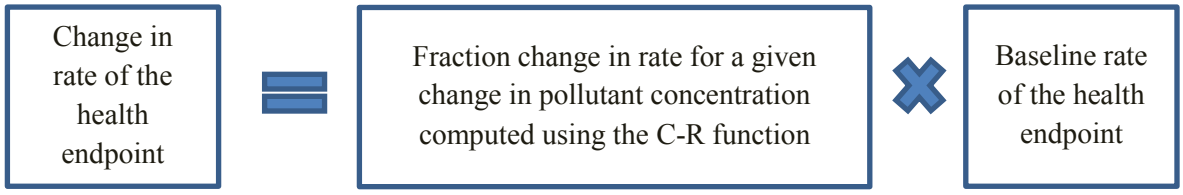
Appendix C. Calculation of the $PM_{2.5}$ and Ozone Components of the Pollution-Vulnerability Index

The Pollution-Vulnerability Index used to identify impacted communities was formed using the sum of three indicators of health impacts from air pollution. The three components used were (1) estimates of increased mortality due to $PM_{2.5}$ and ozone exposures, (2) estimates of increased morbidity costs due to $PM_{2.5}$ and ozone exposures, and (3) estimates of increased cancer risk from toxic air contaminants. This appendix describes how estimates of increased mortality and morbidity costs were calculated. Methods for calculating cancer risk are described in Appendix D.

Previous studies have shown $PM_{2.5}$ and ozone to be related to a range of health endpoints, such as asthma emergency room visits, cardiovascular hospital admissions, and non-accidental mortality. For these and several other effects, the US EPA has determined that the linkage is causal (See, e.g., US EPA 2011, Table 2.1 for $PM_{2.5}$ and US EPA 2007, Section 3.4 for ozone) and has compiled concentration-response (C-R) functions for these endpoints. A C-R function relates a change in the rate of a health endpoint to a change in concentration of a pollutant. For example, an increase in $PM_{2.5}$ concentration of 2 $\mu g/m^3$ in an area results in an estimated increase of 0.92% in the rate of asthma emergency room (ER) visits in adults. A set of C-R functions was available from EPA's BenMAP application (USEPA 2012). In this study, C-R functions were used to estimate the health burden from $PM_{2.5}$ and ozone in each Bay Area zip code area.

C1. Application of the C-R Function.

Estimates of the amount that $PM_{2.5}$ and ozone affect the rates of health endpoints may be determined through application of the C-R function. The basic form of the C-R function for a given health endpoint is:



The precise mathematical formula for the C-R functions used in this report is:

$$\Delta r = \left(1 - e^{-\beta c}\right) r \qquad \text{(C1)}$$

where r is the baseline rate, Δr is the change in the baseline rate, c is the pollutant concentration above a background level, β is the C-R coefficient from BenMAP. So, $1 - e^{-\beta c}$ is the fraction change in rate for a given change in pollutant concentration. Table C1 lists the β coefficients for $PM_{2.5}$ and ozone for various health endpoints.

C1.1. $PM_{2.5}$ and Ozone Background Levels

In this study, the impacts of $PM_{2.5}$ and ozone were determined for increments in concentration above concentration levels representative of clean air, termed *background concentrations*. The background concentration of $PM_{2.5}$ in the Bay Area, which includes natural sources of fine particles such as sea salt, windblown dust, and naturally occurring secondary organic aerosols, was estimated to be 5 $\mu g/m^3$. For use in equation C1, $PM_{2.5}$ concentrations were formed by subtracting 5 $\mu g/m^3$ from the estimated annual $PM_{2.5}$ concentration. The background concentration of ozone, which includes contributions from natural precursors and from stratospheric transport, was estimated to be 40 ppb. For use in equation C1, annual average of daily 8-hour ozone values, x, were truncated at 40 ppb; that is, 8-hour concentrations were assigned a value of zero when below 40 ppb and a value of x – 40, when above 40.

C1.2. Combining $PM_{2.5}$ and Ozone Health Impacts

Some health endpoints have C-R functions for both $PM_{2.5}$ and ozone. Let β_p be the C-R coefficient for $PM_{2.5}$ and β_o be the C-R coefficient for ozone. If the $PM_{2.5}$ and ozone concentrations above background levels for an area are c_o and c_p. Then the combined C-R function is:

$$1 - e^{-(\beta_p c_p + \beta_o c_o)}$$

C1.3. Use of Age Ranges

In general, the C-R functions are specific to only to a particular population age range. We divided ages into ten ranges to take this into account: 0-4, 5-17, 18-24, 25-34, 35-44, 45-54, 55-

64, 65-74, 75-84, and 85+. These correspond to most of the age categories for the C-R functions used here.[5]

C1.4. Pooling C-R Coefficients

For some health endpoints and age ranges, BenMAP lists β effect coefficients for more than one study. For example, for COPD among those 65 years and older, there are two studies, one from Detroit with a coefficient of 0.001169 (standard error = 0.002064; Ito 2003) and one from Los Angeles with a coefficient of 0.001850 (standard error = 0.000524; Moolgavkar 2003). (See Table E-7 in USEPA 2012.)

To generate a single coefficient, we used the "fixed effects weights" method, which weighs coefficients in inverse proportion to their variances. (See Section K.2.1.3 in USEPA 2012.) For the example above, the pooled coefficient β^* for COPD hospital admissions among those 65 years and older was found to be

$$\beta^* = (\beta_1/v_1 + \beta_2/v_2)/(1/v_1 + 1/v_2), =$$

$$(0.001169/v_1 + 0.001850/v_2)/(1/v_1 + 1/v_2) = 0.00181,$$

where v_1 and v_2 are variances with $v_1 = 0.002064^2$, $v_2 = 0.000524^2$.

C2. Estimation and Use of Baseline Rates

As shown in equation C1, the baseline rate is one of the terms in the calculation of the pollution impact of a given health endpoint. The baseline rate is the number of cases of a given endpoint per capita. For example, if the endpoint is asthma hospital admissions, then the baseline rate for a zip code area would be the number of such admissions averaged over three years, i.e., the average annual number, divided by the number of people in the population.

C2.1 Age Adjustment

The raw baseline rate for a given health endpoint is the average annual number of cases of that health endpoint divided by the number of people in the population. In this study we used zip-code specific rates for each age group, but we adjusted the overall rate to correspond to the Bay

[5] The one exception was PM2.5 mortality, where the expert evaluations that produced the C-R coefficient considered the impact on those 30 years and older. In this case, we were able to determine the underlying 30+ mortality rates because we had individual mortality data.

Area age distribution. Thus, if an area has higher rates than the Bay Area average for each age range, then its pollution-vulnerability index will be higher. However, if the area's rates are high because its population is more elderly than the Bay Area age distribution, then the higher rates are corrected for by age adjustment.

Symbolically, for a given endpoint and zip area, we replace the rate of that endpoint, r_a, for a given age category *a*, with an adjusted rate, r_a^*, computed as:

$$r_a^* = r_a * f_a$$

where f_a is the fraction of people in the Bay Area in age category, *a*.[6]

For example, consider the case of the cardiovascular hospital admission endpoint in zip code 94070 (San Carlos). From 2009-2011, there were an average of r_a = 26 cardiovascular admissions per year among 55 to 64 year-olds out of 3,959 residents in that age range. So r_{55-64} = 26/3959 = 0.00657 annual cases per capita in the 55-64 age group. The fraction of Bay Area residents in this age category was 0.119. So r^*_{55-64} = 26/3959 * 0.119 = 0.000782.

C3. Estimation of the Change in Rates Caused by Ozone and $PM_{2.5}$

The estimated change in baseline rate Δr resulting from a given pollutant concentration, *c*, is:

$$\Delta r = \sum_{age\ range\ a}(1 - e^{-\gamma_a})r_a^* \qquad \text{(C2)}$$

where $\gamma_a = I_{oa}\beta_o c_o + I_{pa}\beta_p c_p$, with I_{oa} = 1 if the ozone coefficient applies to age range *a*, and 0 if it does not. Similarly, I_{pa} = 1 if the $PM_{2.5}$ coefficient applies to age range *a*, and 0 if it does not.

Continuing the example in Section C2.1, the annual mean $PM_{2.5}$ concentration in zip 94070 was 8.6 μg/m^3, which is 3.6 μg/m^3 above the assumed background of 5 μg/m^3. From Table C1, the C-R coefficient for cardiovascular hospital admissions for adults 18 to 64 was β = 0.0014 (oops). There is no ozone coefficient for cardiovascular hospital admissions listed in the BenMAP

[6] The term r_a represents the rate within the age category, while r_a^* represents the rate as a fraction of the population. Summing r_a^* values over all age ranges results in a rate per capita.

appendices.[7] So, for 55-64 year olds, $\gamma_a = \gamma_{55\text{-}64} = 0 + 1*0.0014*3.6$. Thus, we estimate the annual increase in cardiovascular hospital admissions due to $PM_{2.5}$ for 55 to 64 year olds in this zip code to be:

$$(1 - e^{-0.0014*3.6})0.000782 = 0.00000393 \text{ admissions for 55-64 year-olds/per zip resident}$$

To apply equation (C2) for this example, we first note that for age ranges 0-4 and 5-17, there are neither ozone nor $PM_{2.5}$ coefficients. Therefore, these age ranges add nothing to the sum. For age ranges 18-24, 25-34, 35-44, 45-54, and 55-64, $\beta = 0.0014$ and $\gamma_a = 0.0014*3.6 = 0.00504$. For ages 65-74, 75-84, and 85+, from the table, $\beta = 0.00107$ (just seeing if you were paying attention.) Thus for these age ranges, $\gamma_a = 0.0007*3.6 = 0.00252$. Therefore, the estimated impact of ozone and $PM_{2.5}$ on cardiovascular admissions in zip code 94070 was:

$$\Delta r = 0 + 0 + (1 - e^{-.00504})r^*_{18-24} + (1 - e^{-.00504})r^*_{25-34} + \cdots + (1 - e^{-.00504})r^*_{55-64} + (1 - e^{-.00252})r^*_{65-74} + \cdots + (1 - e^{-.00252})r^*_{85+}$$

Substituting in the values produces

$$\Delta r = 0.0000165 \text{ cardiovascular hospital admissions per year per capita.}$$

In other words, the levels of $PM_{2.5}$ in San Carlos increase the frequency of cardiovascular hospital admissions to its 29,166 residents by 0.0000165*29166 = 0.48/year. Similar calculations were made for every zip code area and every health endpoint in Table C1. For mortality, this was the final product. For morbitity costs were compiled for the different health outcomes as presented below.

C3.1 Morbidity Costs

For morbidity, because there were many health outcomes, we reduced the number of variables by combining the many outcomes into a single value based on total cost associated with all the outcomes for the zip code area. We use the estimated per-case cost, C_e, for each health endpoint e, where the costs are shown in Table C1 (BenMAP 2012). To the extent possible, the costs

[7] Note this does not necessarily imply there is no effect. It means that the effect size has not been established firmly enough to determine a C-R function.

reflect estimated total social cost (e.g., lost wages, impact on family) as opposed to the cost of health care only. The pollutant morbidity cost was computed as $\sum_{health\ endpoints\ e} C_e \Delta r_e$.

C3.2 More Stable Estimates – Bayes Estimators

The small populations in some zip code areas and the further disaggregation into 10 age categories means that the estimated event rates may be quite variable – subject to chance fluctuations. Several methods were introduced to produce more stable estimators. We've used three years of data rather than just one, and have merged the smallest zips into larger areas. These approaches have been used by others (Kulkami *et al.* 2011,Al-Delaimy *et al.* 2010).

Another method is to use Bayes estimators, namely to use prior information to anchor the estimates. Roughly speaking, Bayes estimators are a weighted average of the raw estimate and a prior estimate. We apply this procedure to the zip area event rates.

Specifically, we assume that the events follow a binomial distribution, and use a beta prior. Here we pretend that event rate is a probability, that is, no individual can have multiple events. This is certainly true for mortality, but there are individuals who might have had multiple trips to the hospital for cardiovascular problems, for example. Details of the method follow below.

Suppose that for a particular zip area and age category there were x cases of a health endpoint (e.g. $x = 10$ hospital admissions for 35-44 year-olds for respiratory disease) and a population of n individuals in that zip area age category. Let x_0 be the number of cases in that age category in the Bay Area as a whole, and n_0 the number of individuals. We assume a prior beta with parameters $\alpha = \gamma x_0$ and $\beta = \gamma(n_0-x_0)$. The posterior distribution is also beta with parameters $\alpha = x + \tau x_0$ and $\beta = n-x + \tau(n_0-x_0)$. The beta distribution is known as a conjugate prior for the binomial – one where the math works out easily. (For more detail, see http://en.wikipedia.org/wiki/Conjugate_prior#Discrete_distributions .)

The parameter τ represents the weight given to the prior distribution. It can be thought of roughly as if we had a prior sample of size γn_0. For this study we used $\tau = 1/1000$ for morbidity

events and $\tau = 1/10{,}000$ for mortality. Since the Bay Area population is about 7,000,000, this effectively assumes a prior sample of 7,000 for morbidity and 700 for mortality.

Figure C1 shows how the mortality rate estimates change by applying the Bayes estimate. A few zip code areas show large changes, but for most, there is little difference.

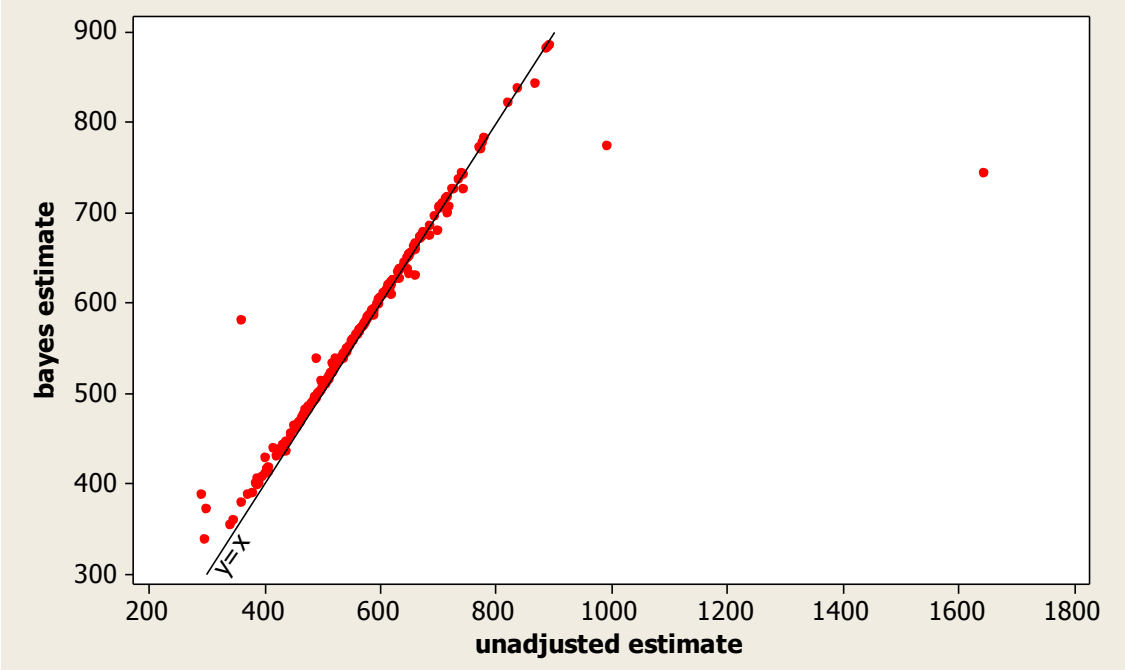


Figure C1. Bayes estimates and unadjusted estimates of non-accidental mortality.

C4. Computing the Pollution-Vulnerability Index

The pollution-vulnerability index is computed from three terms: mortality, morbidity costs and cancer risk. For each term, the zip code areas are ranked from lowest rate to highest. For each zip code area, the three ranks are then summed, and the sums are expressed as a percentage of the largest sum, so that the Pollution-Vulnerability Index ranges between 0 and 100 for all zip code areas.

Table C1. Health Endpoints used in Vulnerability Index – Per case costs, C-R coefficients, and age ranges.

Effects*	$/Case	PMBeta	O3Beta	0-4	5-17	18-24	25-34	35-44	45-54	55-64	65-74	75-84	85+
PM Mortality		0.0100	0	0	0	0	0.5	1	1	1	1	1	1
O3 Mortality		0.0000	0.00049	1	1	1	1	1	1	1	1	1	1
Asthma_HA	12,700	0.0033	0	1	1	1	1	1	1	1	0	0	0
Cardiovascular_HA	32,118	0.0007	0	0	0	0	0	0	0	0	1	1	1
Cardiovascular_HA	34,233	0.0014	0	0	0	1	1	1	1	1	0	0	0
COPD_HA	19,408	0.0018	0.0020	0	0	0	0	0	0	0	1	1	1
COPD_HA	16,758	0.0022	0	0	0	1	1	1	1	1	0	0	0
Myocardial_HA	84,076	0.0026	0	0	0	1	1	1	1	1	1	1	1
Pneumonia_HA	23,097	0.0040	0.0030	0	0	0	0	0	0	0	1	1	1
Respiratory_HA	26,952	0.0021	0.0025	0	0	0	0	0	0	0	1	1	1
Respiratory_HA	12,929	0	0.0082	1	0	0	0	0	0	0	0	0	0
Asthma_ER	468	0.0062	0.0012	1	1	0	0	0	0	0	0	0	0
Asthma_ER	468	0.0046	0.0012	0	0	1	1	1	1	1	1	1	1

*"_HA" = hospital admissions; "_ER" = emergency room visits.

Source: BenMAP Appendices: Tables E-6, E-7, and E-8 for $PM_{2.5}$ and F-3 and F-4 for ozone (BenMAP 2012)

References

Al-Delaimy WK, White MM, Mills AL, Pierce JP, Emory K, Boman M, Smith J, Edland S. Final Summary Report of: *Two Decades of the California Tobacco Control Program: California Tobacco Survey, 1990-2008*, La Jolla, CA: University of California, San Diego; 2010.

Ito, K. 2003. Associations of Particulate Matter Components with Daily Mortality and Morbidity in Detroit, Michigan. In: Revised Analyses of Time-Series Studies of Air Pollution and Health. Health Effects Institute. Boston, MA. May.

Kulkami, SC, A Levin-Rector, M Ezzati, and CJL Murray. "Falling behind: life expectancy in US counties from 2000 to 2007 in an international context." Population Health Metrics 2011, 9:16.

Moolgavkar, S. H. 2003. Air Pollution and Daily Deaths and Hospital Admissions in Los Angeles and Cook Counties. In: Revised Analyses of Time-Series Studies of Air Pollution and Health. Health Effects Institute. Boston, MA. May.

US EPA 2007 Review of the National Ambient Air Quality Standards for Ozone: Policy Assessment of Scientific and Technical Information. OAQPS Staff Paper. EPA-452/R-07-007, July 2007

US EPA 2011 Policy Assessment for the Review of the Particulate Matter National Ambient Air Quality Standards, EPA 452/R-11-003, April 2011

USEPA 2012 *BENMAP Users Manual Appendices*, October, 2012. Prepared for the Office of Air Quality Planning and Standards, USEPA by Abt Associates, In

Appendix D. Cancer Risk Characterization Methods

Excess lifetime cancer risks are estimated as the incremental probability that an individual will develop cancer over a lifetime as a direct result of exposure to potential air pollution carcinogens. The estimated risk is a probability, often expressed as the number of people who might get cancer per million people similarly exposed. The cancer risk attributed to a chemical was calculated over an assumed 70-year lifetime exposure by multiplying by the concentration of each toxic compound by its unit risk factor (UR).[1] A year-specific age sensitivity factor (ASF) increases the risk in early years of exposure to account for increased sensitivities during fetal development and early childhood.[2]

The potential excess lifetime cancer risk was calculated for each individual pollutant, *i*, for each of the years of exposure, *j*:

$$Risk_i = UF_i \times \sum_j^{70\ years} C_{i,j} \times ASF_j,$$

where

$Risk_i$	=	Cancer risk; the incremental probability of an individual developing cancer as a result of inhalation exposure to a particular potential carcinogen *i* (unitless)
UF_i	=	Unit risk factor for pollutant *i* $(\mu g/m^3)^{-1}$
$C_{i,j}$	=	Annual average concentration for pollutant *i* during year *j* $(\mu g/m^3)$
ASF_j	=	Age Sensitivity Factor for year *j*; the value of the factor is higher in early years of exposure (unitless)

In general concentrations vary by year in response to annual average emissions for the year. For this calculation, we assumed that emissions would remain constant for the 70 years of exposure. This is likely a conservative assumption, since emissions toxic air contaminants have been decreasing rapidly over the past several decades. With the assumption of constant emissions, the concentration Ci is constant for each year and the equation for cancer risk from each pollutant is simplified:

$$Risk_i = UF_i \times C_i \times \sum_j^{70\ years} ASF_j \sim UF_i \times C_i \times 1.7\ ,$$

[1] Cal/EPA. 2009. Technical Support Document for Cancer Potency Factors: Methodologies for Derivation, Listing of Available Values, and Adjustment to Allow for Early Life Stage Exposures. May 2009.

[2] Cal/EPA. 2011. OEHHA/CARB Consolidated Table of Approved Risk Assessment Health Values. February 14, 2011. Online: http://www.arb.ca.gov/toxics/healthval/contable.pdf

Since the sum of ASF values for each year is a constant (about 1.7), it does not affect the relative ranking of risk in different communities. For our purposes, we evaluated risk without the constant ASF multiplier.

The total risk is simply the sum of the risk from each pollutant. Table D1 lists the pollutants considered and the unit risk factor. These pollutants represent the top contributors to risk from measurements at the Air Districts toxic monitoring locations.

Table D1. Unit risk factors for modeled pollutants.

Pollutant	Unit Risk Factor (0.001 x *IF* x *CPF*)
Diesel particulate matter	300
1,3-butadiene	170
benzene	29
formaldehyde	6
acetaldehyde	2.7

Cancer risk from five toxics species (diesel PM, 1,3-butadiene, benzene, formaldehyde, and acetaldehyde) was calculated for each grid cell over the entire modeling domain. Other modeled carcinogenic toxics species were not included in this calculation because of their 7 lower concentrations and smaller unit risk factors. The unit risk factors for the above species were 300, 170, 29, 6, and 2.7, respectively. They are expressed as expected excess cancer cases per million per $\mu g/m^3$ These risk values assume a 70-year lifetime exposure.

Cancer risk for each species above was calculated by multiplying its respective annual average concentrations with the corresponding unit risk factor, and then the resulting values were summed across species. The results were expressed as the number of expected cancer incidents per million people.