Supplemental Material

Estimating Risk from Ambient Concentrations of Acrolein across the United States

Concentration Conversions

The goal of this analysis was to predict effects of ambient acrolein exposure in the United States, through creation of a dose-response model and then application of the model to estimated ambient exposure values. We used an animal study to develop a dose-response model for acrolein exposure and lung function (Costa et al. 1986). Two adjustments were performed on the data to allow estimation of adverse risk in humans based on experimental rat data. We adjusted for the differences between human and rat physiology using a regional gas dosimetry ratio (RGDR) factor. In addition, we adjusted for the differences in continuous versus limited exposures by using two time-weighted average factors (TWAs).

We used the Environmental Protection Agency (EPA)'s method for calculating the rat:human RGDR (U.S. EPA 2003); the calculations are:

$$[1]\ \mathrm{RGDR}_{rat:human} = \frac{VR_{rat} / SA_{rat}}{VR_{human} / SA_{human}} = \frac{0.20/15.0}{20/200} = 0.14$$

In this equation, RGDR = regional gas dosimetry ratio (a unitless conversion factor), VR = ventilation rate in m^3/day, and SA = extrathoracic surface area in cm^2; we used EPA estimates for these values (U.S. EPA 2003). As acrolein is a highly reactive Category I gas, the upper respiratory tract is considered the primary target tissue for inhaled acrolein (U.S. EPA 2003).

Exposures in Costa et al. (1986) were subchronic and noncontinuous; therefore, TWA factors were used to convert these to equivalent continuous, chronic exposures as estimated in

the NATA dataset. Although chronic exposure is considered to be any length of time greater than 90 days, we adjusted to annual exposure concentrations because the NATA ambient exposure data is reported as an annual average. The TWA factor consisted of a components to adjust for the exposure continuity [(6/24 hours)* (5/7days)], and duration [(62/365 days)]. We adjusted for the differences between the continuous human exposure concentration and the limited rat exposure concentration using the following equation:

$$[2]\ ACR_{human} = ACR_{rat} * RGDR_{rat:human} * TWA_1 * TWA_2$$

Here, ACR = acrolein concentration in $\mu g/m^3$, $RGDR_{rat:human}$ =0.14, TWA_1 = [(6 hrs / 24 hrs) / (5 days / 7 days)] = 0.18 for noncontinuous:continuous exposures, and TWA_2= [(62 days / 365 days)] = 0.17 for subchronic:chronic exposures. RGDR, TWA_1 and TWA_2 are all unitless conversion factors.

Model Selection

We created models for the effect of various ambient acrolein concentrations on the lung function measurements of specific compliance (sC_L) and the ratio of residual volume to total lung capacity (RV/TLC). Using individual rat data from Costa et al. (1986), human equivalent concentrations (HEC) for acrolein were modeled separately with sC_L and RV/TLC in U.S. EPA's Benchmark Dose Software, version 1.3.2, for three continuous data models: linear, polynomial and power (see Figure 2) (Costa et al. 1986). A likelihood ratio test was used to select between constant or non-constant variance to be used in the models. We used several criteria to select a final model. The critical value of alpha=0.1 was used to assess the fit of the fitted model compared to a fuller model (i.e., a model with individual terms describing each dose group). Chi-square residuals were used to determine which models described the data best at low doses,

likelihood ratio tests were used to evaluate model fit, and values of Akaike's Information Criterion (AIC) were used to rank similar models (U.S. EPA 2000).

Summaries of the fitted models are presented in Table S1. The different models for each of the two lung parameters were roughly similar. All models provided adequate fit in the lower dose ranges and were adequate fits to the overall data. Based on likelihood ratio tests, we used a constant variance model for sC_L and a non-constant variance model for RV/TLC. Upon evaluation of model fit tests (Table S1) and visual inspection of the model graphs, we selected a linear model for sC_L and a power model for RV/TLC. Results from a sensitivity analysis, where sCL was modeled without data from the highest animal dose group (4.0 ppm), are also presented.

In modeling continuous data, BMDS assumes that the data being modeled are approximately normal. The fact that the standard deviation of the response at each dose group is statistically the same suggests that the data are not log-normally distributed (Gaylor and Slikker 1990). We confirmed that the assumption of normality is reasonable based on diagnostic measures including the use of quantile-quantile plots.

Example Calculation of Additional Adverse Outcome

All additional adverse outcome results are presented in Table S2. For this example, we will calculate the additional adverse outcome in terms of the residual volume/total lung capacity ratio (RV/TLC) for the 95^{th} percentile annual acrolein concentration in all counties across the United States in 1999, which was 0.407 $\mu g/m^3$, defining adverse outcome as at or higher than the 90^{th} percentile of the control distribution. Graphic representations of the terms used in the calculations are presented in Figure S1.

As described in the main text, we use the equation $k=z_p-z_{(p-r)}$, adapted from prior work (Crump 1995; Gaylor and Slikker 1990), to compare the change in distributions between the unexposed and exposed populations using a standard normal distribution curve. We are interested in determining *r*, which is the percent of the distribution that has been "shifted" into adverse by the exposure. On a standard normal curve, *k*, z_p and $z_{(p-r)}$ are all values on the x-axis scale; these are standard normal deviates which correspond to a certain percentile (*p*, or *p-r*) of the distribution. The standard normal deviate for the 90th percentile of the population (individuals above this percentile are affected at baseline exposures) is z_p (p=0.90), or 1.28155, and the standard normal deviate for the 0.90-r^{th} percentile (individuals above this are affected at elevated exposure levels) is $z_{(p-r)}$. Thus, *r* is the difference in the proportion of the population affected at baseline exposures and the proportion affected at elevated exposures; *r* represents the estimated additional adverse outcomes from the exposure.

The two standard deviates are related through the term *k*, which represents the difference between the two *z* values. The term *k* is also the multiple of the standard deviation, that is, it is the value multiplied by the standard deviate which represents the change in mean response associated with exposure. In other words, *k* is the difference between the baseline response and the exposed response after standardizing to the standard normal curve. The first step in the process to determine the value of *r* is to determine the value for *k*. To find *k*, we subtract the response at mean exposure from the response at baseline (or in this case no exposure) and divide by the standard deviation. In this example, we predict that mean response at elevated exposure (0.407 μg/m^3) is 0.136047; this is derived from the power model equation from Table S1: $0.136047 = 0.136 + (0.0005(0.407^{2.63}))$. This response transformed into the standard normal scale is (0.136047-0.136) / 0.06304 = 0.000744; where 0.136 is the response without exposure

(baseline) and 0.06304 is the standard deviation of the RV/TLC levels among all animals. The baseline response on the standard normal scale is, by definition, equal to 0 [0.136-0.136) / 0.06304 = 0]. Therefore, the difference between the predicted responses is simply the predicted response at elevated exposure, or 0.000744 (because 0.000744-0=0.000744). So, k=0.000744.

The second step is to determine the value for $z_{(p-r)}$. This is equal to z_p-k, or 1.28155-0.000744, or 1.280806. The cumulative proportion (or area under the curve) for the non-adversely affected proportion of the population associated with 1.280806 is 0.899869: this is the proportion of non-adversely affected individuals in the exposed population. To determine the excess number of adversely affected individuals in this population, we subtract this value from the proportion of non-adversely affected individuals expected in the control population, 0.90. As 0.90-0.899869= 0.000131, we determine that the excess number of adversely affected individuals among the exposed is 0.131 per 1000 (see Table S2).

References

Costa DL, Kutzman RS, Lehmann JR, Drew RT. 1986. Altered lung function and structure in the rat after subchronic exposure to acrolein. Am Rev Respir Dis 133:286-291.

Crump KS. 1995. Calculation of benchmark doses from continuous data. Risk Anal 15:79-89.

Gaylor DW, Slikker W, Jr. 1990. Risk assessment for neurotoxic effects. Neurotoxicology 11:211-218.

U.S. EPA. 2000. Benchmark dose technical guidance document. External review draft. Washington, D.C.:U.S. Environmental Protection Agency, Risk Assessment Forum.

U.S. EPA. 2003. Toxicological Review of Acrolein. Washington, DC: U.S. Environmental Protection Agency.

Table S1: Summary of fitted models[a]

Response	Model	Model Equation[b]	Model Fit[c]	Low Dose Fit[d]	AIC
sC_L	Linear	$Y(x)=0.080-0.055x$	0.72	0.61,-0.44	-476
	Polynomial	$Y(x)=0.081-0.0065x+0.0001x^2$	0.43	0.62,-0.26	-475
	Power	$Y(x)=0.080-0.0055x^1$	0.20	0.13,-0.096	-472
RV/TLC	Linear	$Y(x)=0.125+0.0127x$	0.072	0.67,-1.41	-360
	Polynomial	$Y(x)=0.134-0.0003x+0.0018x^2$	0.076	0.73,-0.030	-360
	Power	$Y(x)=0.136+0.0005x^{2.63}$	0.080	0.13,0.0002	-360
sC_L[e]	Linear	$Y(x)=0.082-0.0085x$	0.35	0.74,-0.22	-413

Abbreviations: AIC = Akaike's Information Criterion; HEC = human equivalent concentration; sC_L = specific compliance; RV = residual volume; TLC = total lung capacity.

a) HEC-transformed data from Costa et al. (1986) were fit to models using BMDS Version 3.1.2, with constant variance for sC_L and non-constant variance for RV/TLC.
b) $Y(x)$ = response, where sC_L is in cm^3/cmH_2O, and RV/TLC has no units; x= HEC acrolein in $\mu g/m^3$.
c) P-values for likelihood ratio test comparing the fitted model to a model allowing for a precise fit of the mean at each dose level (i.e., a fully saturated model); values less than 0.05 suggest the model is not a good fit to the data.
d) $\chi 2$ results comparing observed vs. predicted data for 0.74 and 2.6 $\mu g/m^3$ HEC-acrolein, respectively; values >2 or <-2 indicate a lack of fit.
e) Fit without animals in the highest acrolein dose group.

Table S2: Sensitivity analysis of excess adverse response from ambient acrolein across the United States, 1999[a]

Response	Percentile	Estimated excess adverse response per 1000, AC_{10} [AC_2, AC_{18}]		
		National Average	Urban Counties	Rural Counties
sC_L	5	0.28 [0.043, 0.42]	0.56 [0.086, 0.84]	0.11 [0.017, 0.17]
	10	0.48 [0.073, 0.72]	0.90 [0.135, 1.3]	0.18 [0.027, 0.26]
	25	1.1 [0.17, 1.6]	1.7 [0.26, 2.5]	0.36 [0.054, 0.53]
	50	2.5 [0.38, 3.7]	3.1 [0.48, 4.6]	0.66 [0.10, 0.99]
	75	4.6 [0.71, 6.9]	5.3 [0.81, 7.8]	1.2 [0.18, 1.7]
	90	8.6 [1.3, 13]	9.9 [1.5, 15]	2.0 [0.30, 3.0]
	95	14 [2.2, 20]	15 [2.4, 22]	3.0 [0.46, 4.5]
RV/TLC	5	NC[b]	NC	NC
	10	NC	NC	NC
	25	NC	0.001 [0, 0.001]	NC
	50	0.002 [0, 0.002]	0.003 [0, 0.004]	NC
	75	0.008 [0.001, 0.012]	0.011 [0.002, 0.016]	NC
	90	0.039 [0.006, 0.057]	0.056 [0.008, 0.083]	0 [0, 0.004]
	95	0.13 [0.02, 0.20]	0.157 [0.024, 0.23]	0 [0, 0.004]
sC_L[c]	5	0.48 [0.072, 0.71]	0.94 [0.14, 1.4]	0.19 [0.028, 0.28]
	10	0.80 [0.12, 1.2]	1.9 [0.23, 2.2]	0.30 [0.45, 0.44]
	25	1.9 [0.28, 2.8]	2.9 [0.44, 4.3]	0.60 [0.091, 0.90]
	50	4.2 [0.65, 6.3]	5.2 [0.81, 7.8]	1.1 [0.17, 1.7]
	75	7.9 [1.2, 12]	9.0 [1.4, 13]	1.9 [0.30, 2.9]
	90	15 [2.3, 22]	17 [2.7, 25]	3.4 [0.52, 5.0]
	95	24 [3.9, 35]	26 [4.2, 38]	5.1 [0.78, 7.5]

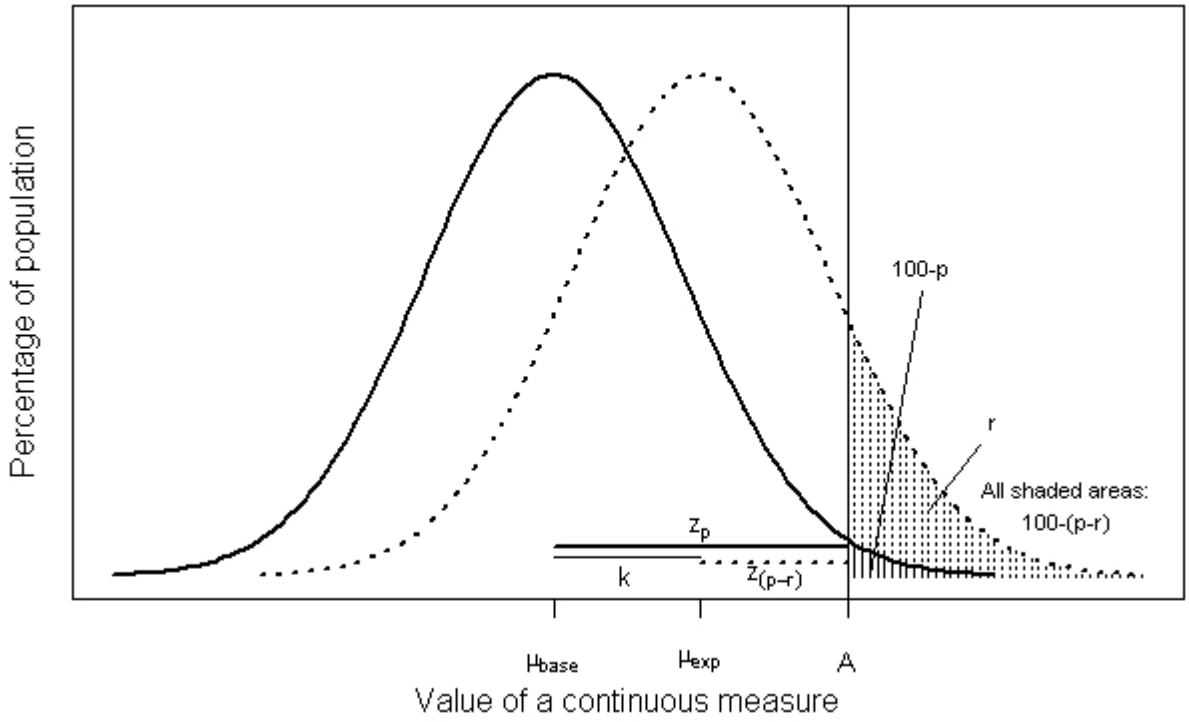
Abbreviations: AC_2, AC_{10}, AC_{18} = adverse cutoff where the baseline prevalence is 2,10, or 18 percent respectively; sC_L = dynamic compliance; RV = residual volume; TLC = total lung capacity.

a) Excess adverse responses are the estimated additional number of adverse responses per 1000 people; an adverse response is a response at or higher than the designated percentile response among unexposed (control) individuals. Estimated excess effects for sC_L at acrolein's reference concentration (0.02 $\mu g/m^3$) are 0.65 [0.099, 0.97].
b) No change from baseline was detected.
c) Estimates from model without the highest dose group.

Figure Legend

Reproduction of Figure 1 (in the main text) on a standard normal scale, illustrating elements used in calculation of additional adverse outcomes. As in Figure 1, the curve with the solid line represents the distribution of response severity among the baseline (unexposed) population and the curve with the dotted line represents the distribution of response severity among the exposed population. On the x-axis, μ_{base} represents the mean baseline (no exposure) response and μ_{exp} represents the average response at an elevated exposure level; A represents the designated cutoff response level above which responses are considered adverse. The three lines above the mean responses represent (from highest to lowest) the values of k, z_p and $z_{(p-r)}$; the relationship between these elements is $z_p - z_{(p-r)} = k$. The area shaded with straight lines represents the proportion of the population with adverse responses at baseline exposure, 100-p; the combination of the areas shaded with lines and dots represents the proportion of the population with adverse responses at elevated exposure, 100-$(p-r)$. Lastly, the area shaded with dots represents the proportion of the population with adverse responses at elevated exposure that did not have an adverse response at the baseline (no exposure) level, or the proportion of additional adverse outcomes, r.

Figure S1


Percentage of population
100-p
r
All shaded areas:
100-(p-r)
z_p
k
$z_{(p-r)}$
μ_{base}
μ_{exp}
A
Value of a continuous measure