

CASAC Ozone Panel: Comments on the Policy Assessment

Richard L. Smith

Department of Statistics and Operations Research,

University of North Carolina, Chapel Hill

rls@email.unc.edu

Oral presentation March 26, 2014; Written version March 31, 2014

I am Richard Smith, Professor of Statistics and Biostatistics at the University of North Carolina, and also a consultant to the American Petroleum Institute. The opinions expressed here are my personal views and not the official policy of UNC or API.

This comment concerns the ozone chamber studies which are discussed in chapter 6 of the HREA and also in chapters 3 and 4 of the Policy Assessment. In a series of papers between 2000 and 2006, Adams studied lung function responses to ozone (principally FEV₁ decrement) under a variety of treatment regimes for healthy adult subjects exercising while exposed to ozone. Adams (2006) found no statistically significant effect due to ozone at 60 ppb using a Scheffe multiple comparisons test, but Brown et al. (2008) did find significant results using alternative testing methods, a result that was apparently upheld in a court case in Mississippi. In previous comments to CASAC (October 9, 2007), I have argued that there is still a need to take account of multiple testing even if the Scheffe approach is rejected as too conservative.

There are now new results due to Schelegle (2009) and Kim (2011) which reinforce the statistical significance of FEV₁ decrement at 70 ppb and 60 ppb respectively, but the latter result due to Kim is very slight – the mean FEV₁ decrement at 60 ppb is only 1.7%, well below the 10% FEV₁ decrement usually considered clinically significant. The PA focusses on a different issue, the proportion of individuals in all four studies with FEV₁ decrement more than 10%, but this proportion is itself only 10% at 60 ppb, a result that could easily be due to unexplained individual variability.

Another variable also considered in these studies is polymorphonuclear neutrophils in sputum (PMN), studied by Alexis et al. (2010) for 80 ppb ozone exposure and by Kim et al. (2011) for 60 ppb ozone exposure. However, as far as I can tell, there is still no agreement of what level of PMN is clinically significant – I can find no discussion of this issue in the present documents. According to personal communications with Dr. Alexis, PMN count is more relevant than percent PMN, but Kim et al. (2011) did not find a statistically significant effect in PMN count, only in percent PMN.

An alternative way to think about the variability in these results is to try to explain it in terms of other measured covariates, using regression analysis. My analysis of the 60 ppb FEV₁ decrement results of Kim et al. led to rather a nice result: a statistically significant relationship between FEV₁ decrement after ozone exposure, and baseline FEV₁, adjusted for the individual's height. This could indicate that individuals who already suffer from poor lung function are more susceptible to ozone exposure than normal individuals. Unfortunately, a parallel analysis based on %PMN did not yield any significant covariates. It remains my impression that these experiments are too small, and their statistical interpretation too limited, for them to have any clear significance in the regulatory context.

Supplementary Analyses

Kim et al. (2011) studied the lung function effects of exposure to ozone at 0.06 ppm for 6.6 hours in healthy exercising young adults. The mean decrement in FEV₁, adjusted by comparison with the corresponding result in clean air, was 1.7%, a result they claimed as statistically significant, though well short of the 10% that is considered by EPA (see e.g. page 6-14 in the Integrated Science Assessment for Ozone and Related Photochemical Oxidants; Second External Review Draft, Environmental Protection Agency, September 2011) to be medically significant. I have re-analyzed the raw data that were kindly provided by Dr. Kim.

I have three main points about the analysis of these data. First, as in all such datasets, there are outliers which are hard to attribute to any medical effect of ozone – one individual exhibited a 17% reduction in FEV₁ in clean air. I do not believe it is valid to over-interpret individual outlier results, for example extrapolating their prevalence to represent population levels of susceptible individuals – rather, the goal of a statistical analysis should be to find overall patterns and trends in the data that are robust against such outliers. The distribution of the CA-adjusted FEV₁ decrements (Figure 1) is too peaked in the middle and too long-tailed to be considered normal – indeed, several standard tests of normality resulted in clear rejection of the null hypothesis. (The tests considered were Looney-Gulledge, Kolmogorov-Smirnov, Cramér-von Mises and Anderson-Darling. Percentage points were computed by simulation allowing for the estimation of mean and standard deviation. For Kolmogorov-Smirnov, the P-value for the null hypothesis that the data are normally distributed was 0.01; for the other three tests, it was substantially smaller.) Because of this, I conclude that tests and confidence intervals based on the t distribution are not valid. Nevertheless, an alternative approach based on the bootstrap does show statistical significance in several of the basic 6.6 hour results (Table 1). To derive this, I computed the standard t statistic, for each bootstrap sample, by subtracting the original sample mean from the bootstrap sample mean and dividing by the bootstrap standard deviation. The sampling distribution of this statistic, from 100,000 bootstrap samples, was then used in place of the t distribution for constructing tests and confidence intervals.

Fig. 1: Histogram for CA-Adjusted FEV1 Decrement

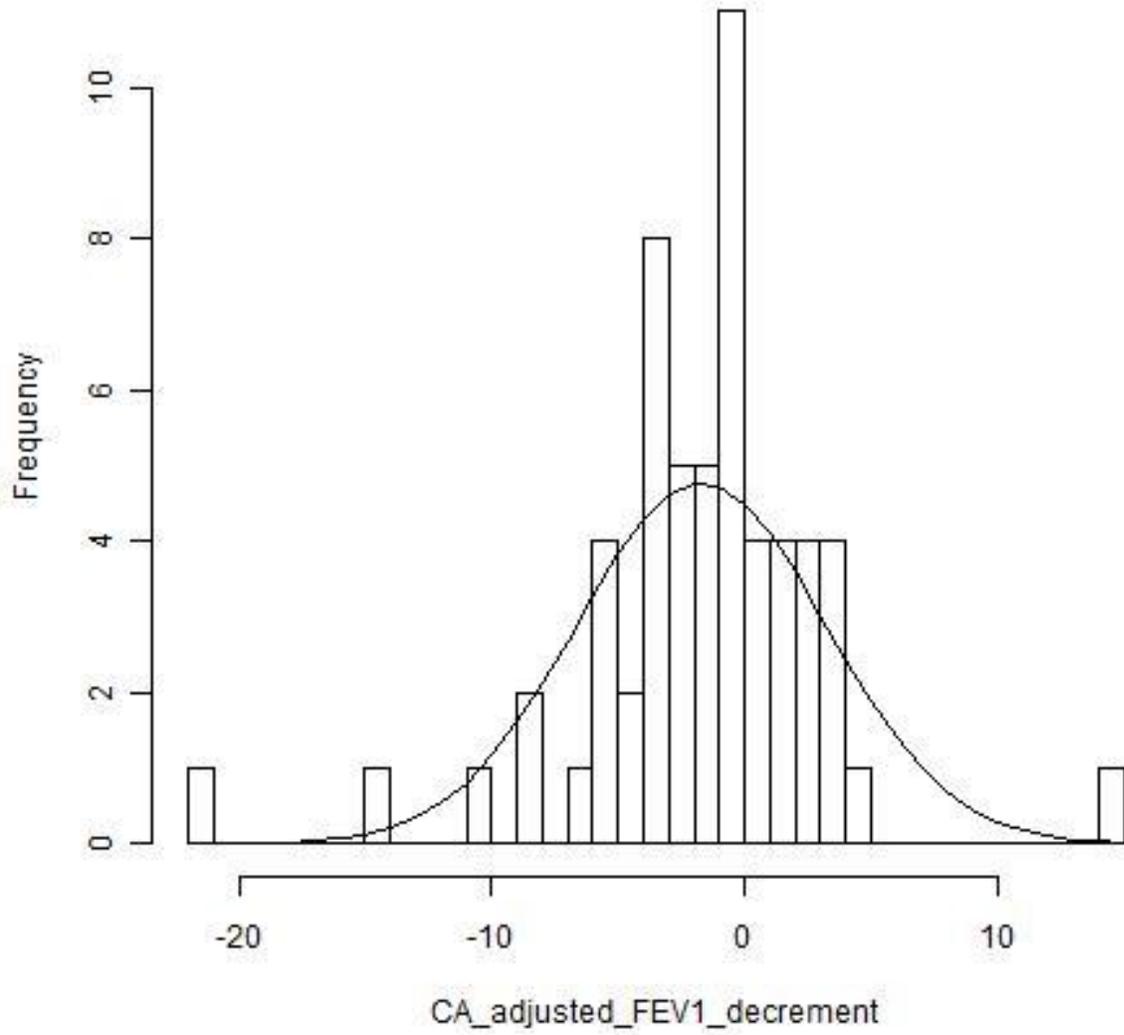


Table 1: Confidence Intervals and P-values For 6.6-Hour Responses

Variable	Sex	N	Estimate	S.E.	This Report Using Bootstrap		Kim (2011) Using <i>t</i> distributions	
					95% CI	P-value	95% CI	P-value
FEV ₁ at 0.06ppm	All	59	-1.71	0.64	(-3.12,-0.54)	0.003	(-3.0,-0.5)	< .05
FEV ₁ at 0.06ppm	Male	27	-1.46	0.96	(-3.22,0.23)	0.086	(-3.4,0.5)	> .05
FEV ₁ at 0.06ppm	Female	32	-1.92	0.88	(-3.92,-0.13)	0.035	(-3.8,-0.3)	< .05
FEV ₁ at 0.08ppm	All	30	-3.38	1.05	(-5.43,-1.13)	0.004		
FEV ₁ at 0.08ppm	Male	15	-4.84	1.46	(-7.83,-2.08)	0.002		
FEV ₁ at 0.08ppm	Female	15	-1.93	1.46	(-4.88,1.52)	0.26		
FVC at 0.06ppm	All	59	-1.19	0.51	(-2.37,-0.29)	0.008	(-2.2,-0.2)	< .05
FVC at 0.06ppm	Male	27	-0.49	0.74	(-1.93,0.88)	0.462	(-2.0,1.0)	> .05
FVC at 0.06ppm	Female	32	-1.78	0.68	(-3.42,-0.48)	0.007	(-3.1,-0.4)	< .05
FVC at 0.08ppm	All	30	-2.36	0.8	(-4.19,-0.87)	0.002		
FVC at 0.08ppm	Male	15	-3.25	1.12	(-6.06,-0.94)	0.007		
FVC at 0.08ppm	Female	15	-1.46	1.12	(-3.56,0.65)	0.165		

My second point, however, is that statistical significance measures lead to variable results when applied to different data samples collected in the experiment. Apart from the data reported in the paper, Kim et al. also recorded intermediate results at 3, 4.6 and 5.6 hours, at 18-hours post-test for a subset of participants, and also at 0.08 ppm ozone exposure for another subset of subjects. The bootstrap results from these experiments (Table 2) show some inconsistencies. For example, in several cases the result is stronger at 0.06 ppm than 0.08 ppm. My conclusion is not to put too much faith in statistical significance: there is still much unexplained experimental variability.

My third point is that it is possible to extend Kim et al.'s results using regression analysis. I regressed the CA-adjusted FEV₁ decrements on several potential explanatory variables. Variables considered were sex, age, height, weight, BSA, minute ventilation per square meter BSA, baseline FEV₁, and an indicator of whether the same individual was also included in the 0.08 ppm experiment. All variables except height and baseline FEV₁ were eliminated by backward selection. Further analysis showed that this could be reduced to a single explanatory variable, baseline FEV₁ adjusted for height, or in other words, the residual when baseline FEV₁ is regressed against height. This variable could be an indicator of prior disease. However, even if the regression line (Figure 2) is interpreted literally, the adjusted baseline FEV₁ would have to be two standard deviations below the mean to produce a 5% predicted FEV₁ decrement, and five standard deviations below the mean for a 10% predicted decrement. Even accepting that FEV₁ measurements are not normally distributed, I doubt that there are many people in the population whose baseline FEV₁ is five standard deviations below its predicted value, and even then, the result relies on extrapolation well beyond the range of the actual data.

Table 2: Results for Other Endpoints

Variable	Time	N	Estimate	S.E.	95% CI	P-value
FEV1 at 0.06ppm	3 hr	59	-0.44	0.42	(-1.27,0.40)	0.303
FEV1 at 0.08ppm	3 hr	30	-0.42	0.63	(-1.65,0.95)	0.53
FEV1 at 0.06ppm	4.6 hr	59	-1.46	0.59	(-2.82,-0.4)	0.007
FEV1 at 0.08ppm	4.6 hr	30	-1.07	0.94	(-2.99,0.83)	0.268
FEV1 at 0.06ppm	5.6 hr	59	-1.50	0.60	(-2.63,-0.24)	0.018
FEV1 at 0.08ppm	5.6 hr	30	-2.57	1.06	(-4.55,-0.19)	0.035
FVC at 0.06ppm	3 hr	59	-0.31	0.40	(-1.07,0.52)	0.449
FVC at 0.08ppm	3 hr	30	-0.28	0.65	(-1.53,1.14)	0.685
FVC at 0.06ppm	4.6 hr	59	-1.15	0.45	(-2.1,-0.29)	0.008
FVC at 0.08ppm	4.6 hr	30	-0.54	0.63	(-1.88,0.7)	0.386
FVC at 0.06ppm	5.6 hr	59	-1.13	0.48	(-2.22,-0.26)	0.01
FVC at 0.08ppm	5.6 hr	30	-2.07	0.78	(-3.8,-0.56)	0.008
FEV1 at 0.06ppm	18-hr post	19	-2.37	0.81	(-4.14,-0.71)	0.007
FEV1 at 0.08ppm	18-hr post	16	-1.45	1.56	(-5.61,1.39)	0.316

To confound things even further, when I attempted the same regression analysis for the experiments with 0.08 ppm ozone, I found no significant covariates at all. In this case, when CA-adjusted FEV₁ decrement following exposure to ozone at 0.08 ppm was regressed against the height-adjusted baseline FEV₁, the estimated slope was 0.57, the standard error 1.05, but the 95% bootstrap confidence interval (-2.87, 3.23) overlapped zero, showing that the result was not significant at 0.05. The corresponding results at 0.06 ppm (Figure 2) were an estimated slope of 1.71, a standard error of 0.61, and a 95% confidence interval (0.2, 3.41) that was significant at P=0.05.

Analysis of %PMN data at 0.06 ppm

Another analysis reported by Kim et al. (2011) was that the difference in %PMN (Ozone-CA) at 0.06 ppm exposure to ozone is 15.7 (standard error: 3.1). Subdivided into men and women, the corresponding numbers were 24.2 (4.3) for men and 8.5 (3.7) for women. All of these are statistically significant at the 0.05 level. The difference between men and women is also statistical significant. However, analysis using GSTM1 as a predictor did not show a statistically significant effect. Dr. Kim has kindly provided me with the individual %PMN numbers used for these comparisons and I succeeded in recomputing Kim's Table 4 based on these data.

X7: Weight

X8: Body surface area

X9: Pre-exposure FEV1 in clean air

X10: Pre-exposure FEV1 in 0.06ppm ozone

The intention behind including the variables X5, X9 and X10 was to see whether variables that might indicate the prior health of the subject had an influence on the change on %PMN due to ozone. The other variables are general “personal characteristic” variables that may be relevant in determining vulnerability. Note that all the subjects were young (age range 20-33), so we would not expect age to have a major influence.

Best subset regression was used to select the best model of each model order, followed by an examination of the selected models to determine which variables were statistically significant. The result was clear-cut: sex is the only significant covariate among the ones listed above. The final fitted model is as shown in the Table 3.

Table 3: ANOVA table for Regression
Dependent Variable: %PMN(Ozone-CA) at 0.06 ppm

Residuals:

Min	1Q	Median	3Q	Max
-33.782	-10.410	1.417	8.415	26.218

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	8.485	3.719	2.282	0.03254 *
Sex	15.697	5.493	2.858	0.00915 **

Signif. codes: 0 *** 0.001 ** 0.01 * 0.05 . 0.1 1

Residual standard error: 13.41 on 22 degrees of freedom

Multiple R-squared: 0.2707, Adjusted R-squared: 0.2376

F-statistic: 8.167 on 1 and 22 DF, p-value: 0.009147

This essentially confirms the result of Kim et al.’s Table 4 – the only statistically significant variable was sex, with an estimated coefficient (male minus female response) of 15.5, standard error 5.5, significant at $p < 0.01$.

Analysis of %PMN data at 0.08 ppm

Now let us consider the data of Alexis et al. (2010). In this experiment, there was only one ozone level of 0.08 ppm, with no control experiment in clean air, a point which the authors acknowledged was a deficiency of their study design. In place of a clean air measurement, the authors measured %PMN and a second variable, total cell count (in Cells/mg), both before and after the experiment. These two variables measure different things with cell count being possibly the better measure of impact (Dr. Neil Alexis, personal communication) but it is also subject to more variability; indeed, for the experiment at 0.06 ppm, the difference in total cell count was not statistically significant as reported by Kim et al. (2011).

For the present analysis, I have repeated the same form of analysis as was done with the %PMN data at 0.06 ppm but using the pre-exposure value of %PMN as the control variable. Note that, for this result to be comparable with the previously reported results in 0.06 ppm ozone, we would effectively be assuming that there is no difference in %PMN in clean air due to exercise alone; this is not certain but is very likely correct (Dr. Neil Alexis, personal communication).

The variables used in the regression in this case were:

X1: Pre-exposure %PMN

X2: sex (=1 for male, =0 for female)

X3: age

X4: Height

X5: Weight

X6: Body surface area

X7: Pre-exposure FEV1 in clean air

X8: Pre-exposure FEV1 in 0.06ppm ozone

X9: Pre-exposure FEV1 in 0.08ppm ozone

The significant variables in this case were X1, X8 and X9, which are all pre-exposure measures. The ANOVA table and related statistics in this case are in Table 4.

Analysis of Total Cell Count at 0.08 ppm

In this case we took the logarithm of total cell count as the variable of interest, since the data are highly right-skewed and taking logarithms give a closer fit to the normal distribution. By analogy with the %PMN analysis, the difference (post-exposure minus pre-exposure) in log total cell count was taken as the dependent variable in a linear regression, while the variable X1 in the

%PMN analysis was replaced by the pre-exposure total cell count. The optimal model in this case is as shown in Table 5. Here, X3, X4 and X7 were the significant variables.

Table 4: ANOVA table for Regression
Dependent Variable: %PMN(Post-Pre) at 0.08 ppm

Residuals:

Min	1Q	Median	3Q	Max
-14.0944	-6.6883	-0.5837	3.3436	22.7282

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	72.4755	19.1605	3.783	0.00303 **
Pre-exp %PMN	-0.5933	0.1609	-3.687	0.00358 **
Pre-exp FEV1 at 0.06ppm	60.4067	24.0749	2.509	0.02903 *
Pre-exp FEV1 at 0.08ppm	-67.9447	23.1347	-2.937	0.01352 *

Signif. codes: 0 *** 0.001 ** 0.01 * 0.05 . 0.1 1

Residual standard error: 11.14 on 11 degrees of freedom

Multiple R-squared: 0.6529, Adjusted R-squared: 0.5583

F-statistic: 6.898 on 3 and 11 DF, p-value: 0.007032

Differences in %PMN values between 0.06ppm and 0.08ppm

Now we turn to what may possibly be the most critical of the various statistical analyses, which is the comparison between the results at 0.06 ppm ozone and 0.08 ppm ozone. As noted already, the two experiments are not strictly comparable because of the different control measurements, but there is a actually very little evidence that they are different. Note that the two experiments were based on distinct groups of subjects, so the comparison is of the “two-sample t-test” type, not a paired comparison.

For pre-exposure %PMN in the 0.08 ppm experiment, the mean was 36.6 and the standard error 5.00

For the post-exposure %PMN in clean air, that was part of the 0.06 ppm experiment, the mean was 38.3 and the standard error 3.71.

For the difference, the mean was 1.70 and the standard error 6.23 (t=0.27, clearly not significant).

Table 5: ANOVA table for Regression
Dependent Variable: Log total PMN cell count (Post-Pre) at 0.08 ppm

Residuals:

Min	1Q	Median	3Q	Max
-0.9920	-0.3572	-0.1253	0.2635	1.2611

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-16.14098	7.35641	-2.194	0.05060 .
Age	-0.13801	0.04906	-2.813	0.01688 *
Height	0.15903	0.05479	2.902	0.01438 *
Pre-exp FEV1 in CA	-1.71558	0.51494	-3.332	0.00669 **

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.6964 on 11 degrees of freedom
Multiple R-squared: 0.5732, Adjusted R-squared: 0.4568
F-statistic: 4.924 on 3 and 11 DF, p-value: 0.02086

This confirms that there is no difference (in this analysis) between the pre-exposure reading at 0.08 ppm and the clear air reading with the 0.06 ppm cohort.

From now on, we assume that there is no difference between these two control measurements, and combine the two sets of data in a single regression analysis. The available covariates for this are

X1: Control level of %PMN

X2: Ozone concentration (coded 0 for 0.06 ppm, 1 for 0.08 ppm)

X3: sex (=1 for male, =0 for female)

X4: age

X5: Height

X6: Weight

X7: Body surface area

X8: Pre-exposure FEV1 in clean air

X9: Pre-exposure FEV1 in 0.06ppm ozone

The variable X9 needs some explanation. It appears that the subjects who participated in the Alexis et al. (2010) experiment at 0.08 ppm also participated in the Kim et al. (2011) experiment

at 0.06 ppm This is why X9 is available for subjects in both experiments. However, it seems that sputum measurements were only taken for the new subjects, not the ones who participated in the earlier experiment. A better comparison between the 0.06 ppm and 0.08 ppm ozone experiments could have been made if the measurements were repeated on the same subjects.

Nevertheless, I have conducted another linear regression analysis using the data as available. Again, the regression strategy was to use all-subsets regression to determine the best model of each order, following by checking the individual regression models for statistical significance of the coefficients. In this analysis, none of the covariates in any of the regression analysis (except for the intercept) was statistically significant at the 0.05 level. As an illustration, Table 6 shows the analysis with just ozone level as covariate.

Table 6: ANOVA table for Regression
Dependent Variable: %PMN(Post-Pre), Combined Experiment

```

Residuals:
    Min       1Q   Median       3Q      Max
-27.957 -13.212   1.121  11.207  34.721

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)   15.679      3.246   4.830 2.38e-05 ***
Ozone level    6.198      5.234   1.184  0.244
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 15.9 on 37 degrees of freedom
Multiple R-squared:  0.03652,    Adjusted R-squared:  0.01048
F-statistic: 1.402 on 1 and 37 DF,  p-value: 0.2439

```

The result confirms the statistical significance of the intercept, which represents the average rise in %PMN over all 39 subjects in the two experiments (15.7 with a standard error of 3.2). However, the difference between the two ozone levels (represented by X2) is not statistically significant. As a result, it is not possible to confirm a dose-response effect for this experiment.

References

Adams WC. Comparison of chamber 6.6-h exposures to 0.04–0.08 ppm ozone via square-wave and triangular profiles on pulmonary responses. *Inhal Toxicol* 2006;18:127–136.

Alexis NE, Lay JC, Hazucha M, Harris B, Hernandez ML, Bromberg PA, Kehrl H, Diaz-Sanchez D, Kim C, Devlin RB and Peden, DB (2010), Low-level exposure induces airways

inflammation and modifies cell surface phenotypes in healthy humans. *Inhalation Toxicology* 22(7), 593-600.

Brown JS, Bateson TF and McDonnell WF (2008), Effects of exposure to 0.06 ppm ozone on FEV1 in humans: A secondary analysis of existing data, *Environmental Health Perspectives* 116, 1023-1026

Kim CS, Alexis NE, Rappold AG, Kehrl H, Hazucha MJ, Lay JC, Schmitt MT, Case M, Devlin RB, Peden DB and Diaz-Sanchez D (2011), Lung Function and inflammatory responses in healthy young adults exposed to 0.06 ppm ozone for 6.6 Hours, *American Journal of Respiratory and Critical Care Medicine* 183, 1215-1221.

Schelegle ES, Morales CA, Walby WF, Marion S, Allen RP. 6.6-Hour inhalation of ozone concentrations from 60 to 87 parts per billion in healthy humans. *Am J Respir Crit Care Med* 2009;180: 265–272.