

## COMMENT FOR CASAC OZONE PANEL – JANUARY 9, 2012

**RICHARD L. SMITH**

I am Professor of Statistics at the University of North Carolina, Chapel Hill. I have been working on air pollution health-related research since about 1995. The following remarks were prepared as a consultant to the American Petroleum Institute but they represent my personal opinions and not the official position of either UNC or API.

The recent paper by Kim et al. (2011) has added to our knowledge<sup>1</sup> of 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 FEV<sub>1</sub>, 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<sup>2</sup> 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 FEV<sub>1</sub> 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 FEV<sub>1</sub> decrements (Figure 1) is too peaked in the middle and too long-tailed to be considered normal – indeed, several standard tests of normality<sup>3</sup> resulted in clear rejection of the null hypothesis. 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<sup>4</sup> does show statistical significance in several of the basic 6.6 hour results (Table 1).

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 FEV<sub>1</sub> decrements on several potential explanatory variables<sup>5</sup> including sex, age, height and baseline FEV<sub>1</sub>. The one statistically significant explanatory variable was baseline FEV<sub>1</sub> adjusted for height<sup>6</sup>, a variable that could be an indicator of prior disease. However, even if the regression line (Figure 2) is interpreted literally, the adjusted baseline FEV<sub>1</sub> would have to be two standard deviations below the mean to produce a 5% predicted FEV<sub>1</sub> decrement, and five standard deviations below the mean for a 10% predicted decrement. Even accepting that FEV<sub>1</sub> measurements are not normally distributed, I doubt that there are many people in the population whose baseline FEV<sub>1</sub> is five standard deviations below its predicted value, and even then, the result relies on extrapolation well beyond the range of the actual data.

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<sup>7</sup>.

In conclusion, the recent paper by Kim et al. contains evidence of a small but statistically significant response to 0.06 ppm ozone at 6.6 hours. However, analysis of the same data at different time points and comparing results at 0.06 and 0.08 ppm ozone still does not show a consistent pattern of responses. To find evidence in this dataset for medically significant effects, even in a susceptible subset of the population, requires considerable extrapolation beyond the range of the data, of the kind that statisticians interpreting experimental data typically warn *against*.

Reference:

Kim, C.S. *et al.* (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.

**Fig. 1: Histogram for CA-Adjusted FEV1 Decrement**

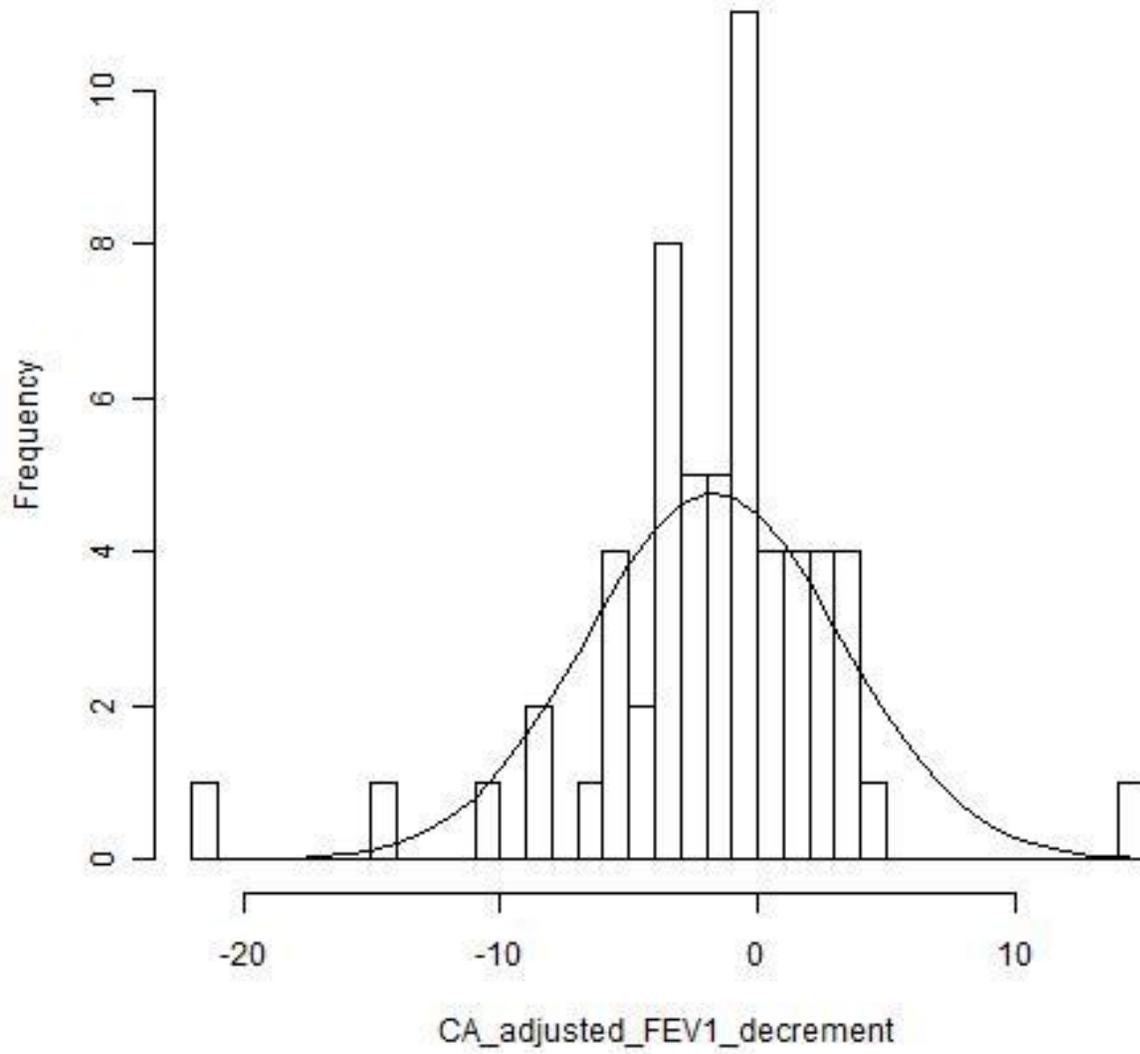
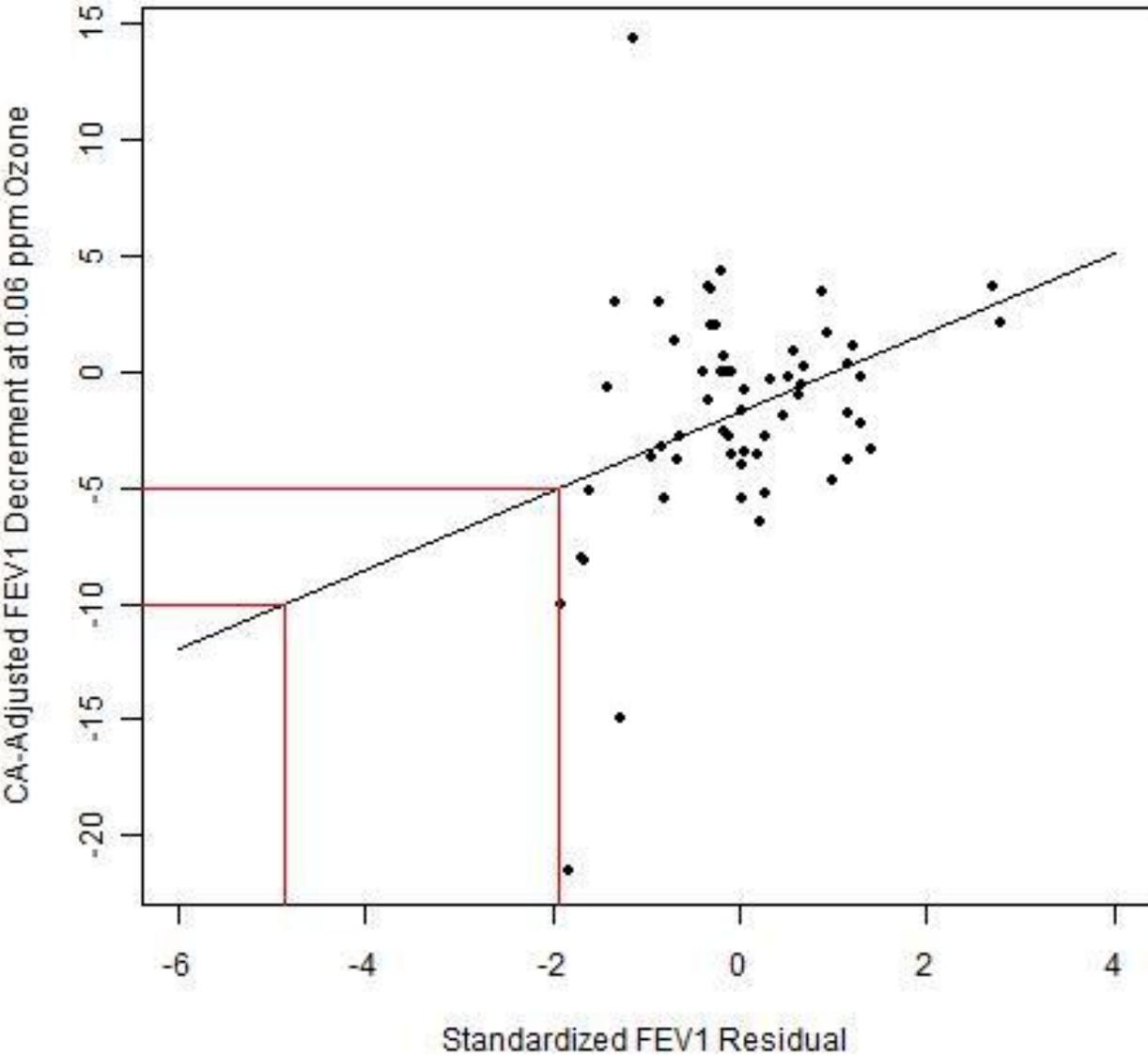


Figure 2: Regression Results



**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 <sub>1</sub> at 0.06ppm	All	59	-1.71	0.64	(-3.12,-0.54)	0.003	(-3.0,-0.5)	< .05
FEV <sub>1</sub> at 0.06ppm	Male	27	-1.46	0.96	(-3.22,0.23)	0.086	(-3.4,0.5)	> .05
FEV <sub>1</sub> at 0.06ppm	Female	32	-1.92	0.88	(-3.92,-0.13)	0.035	(-3.8,-0.3)	< .05
FEV <sub>1</sub> at 0.08ppm	All	30	-3.38	1.05	(-5.43,-1.13)	0.004		
FEV <sub>1</sub> at 0.08ppm	Male	15	-4.84	1.46	(-7.83,-2.08)	0.002		
FEV <sub>1</sub> 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		

**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

## Endnotes

1. Previous papers examining 0.06 ppm ozone include Adams, W.C. (2006), Comparison of chamber 6.6-h exposures to 0.04-0.08 ppm ozone via square-wave and triangular profiles on pulmonary responses. *Inhalation Toxicology* **18**(2), 127-136; Brown, J.S., Bateson, T.F. and McDonnell, W.F. (2008), Effects of exposure to 0.06 ppm ozone on FEV<sub>1</sub> in humans: A secondary analysis of existing data, *Environmental Health Perspectives* **116**, 1023-1026; Nicolich, M. (2007), Some additional statistical analyses of the FEV<sub>1</sub> pulmonary response data from the W.C. Adams data (2006), Attachment A to the October 9, 2007 Exxon comments on the 2008 O<sub>3</sub> NAAQS (<http://www.regulations.gov/#!home:docket#EPA-HQ-OAR-2005-0172-4163>); Schelegle, E.S. *et al.* (2009), 6.6-Hour inhalation of ozone concentrations from 60 to 87 parts per billion in health humans, *American Journal of Respiratory and Critical Care Medicine* **180**, 265-272.
2. See discussion on page 6-14 in the Integrated Science Assessment for Ozone and Related Photochemical Oxidants; Second External Review Draft, Environmental Protection Agency, September 2011.
3. The tests applied were Looney-Gulledge, Kolmogorov-Smirnov, Cramér-von Mises and Anderson-Darling. In all cases the 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.
4. 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.
5. Variables considered were sex, age, height, weight, BSA, minute ventilation per square meter BSA, baseline FEV<sub>1</sub>, and an indicator of whether the same individual was also included in the 0.08 ppm experiment. All variables except height and baseline FEV<sub>1</sub> were eliminated by backward selection.
6. The final regression variable used was the residual when baseline FEV<sub>1</sub> is regressed against height.
7. When CA-adjusted FEV<sub>1</sub> decrement following exposure to ozone at 0.08 ppm was regressed against the height-adjusted baseline FEV<sub>1</sub>, the estimated slope was 0.57, the standard error 1.05, but the 95% bootstrap confidence interval (-2.87, 3.23) 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.