PUBLIC COMMENT TO CASAC OZONE REVIEW PANEL MARCH 5 2007

Richard L. Smith
Department of Statistics and Operations Research
University of North Carolina
Chapel Hill, NC 27599-3260
rls@email.unc.edu

I am a professor of statistics at the University of North Carolina, Chapel Hill. My research interests include the epidemiology of particulate matter and ozone. My recent research sponsors include NSF, NOAA, NIEHS and the American Petroleum Institute. However the comments that follow are my own personal opinions and do not represent the views of the University of North Carolina or any of my research sponsors.

My comments relate to the ozone chamber experiments of Adams (2006) and earlier related papers. These experiments have characterized the effect on lung function (measured by FEV_1 decrement) of 6.6-hour exposure to ozone at a variety of levels (filtered air, 0.04 ppm, 0.06 ppm, 0.08 ppm, ...) and two concentration patterns (square and triangular) (see pages 3–6 to 3–9 and 5–18 to 5–28 of the staff paper). I have two major comments to make:

Point 1. Adams (2006) found a significant effect of ozone on FEV_1 decrement at 0.08 ppm ozone, but not at 0.06 ppm. The staff paper criticizes Adams' use of a Scheffé multiple comparisons procedure and provides an alternative analysis that does show an effect at 0.06 ppm. However, Adams' analysis was designed to protect against possibly spurious effects being detected when comparing many experiments simultaneously. When this aspect is taken into account, the evidence for a response at 0.06 ppm ozone level is still very uncertain.

The attached table shows all possible pairwise t-test comparisons among the Adams experiments (including one from Adams (2002), using a different cohort, for the 0.04 ppm square wave experiment). The table shows the mean decrement in FEV_1 for each comparison, with asterisks denoting the level of statistical significance.

In making policy-relevant comparisons, those with 0.04 ppm ozone level are more relevant than those with filtered air, which does not represent a realistic background level. However, of the four possible comparisons between 0.04 ppm and 0.06 ppm, only two of them are statistically significant, and the mean FEV_1 decrement is less than half the corresponding mean at 0.08 ppm ozone level.

In my judgment, when all the comparisons are taken into account, there is insufficient evidence to conclude that there is any well-defined response to ozone exposure below the 0.08 ppm level.

Point 2. The second type of comparison based on the Adams experiments is for the proportion of individuals who show an FEV_1 decrement greater than 10%. See in particular, Fig. 5.2(a) on page 5–24, which plots the dose-response curve and associated confidence limits. From visual inspection of that curve, it looks as though a 95% confidence interval for the response at 0.06 ppm ozone is between roughly 6% and 16%.

However, the actual experiment shows that at 0.06 ppm ozone exposure (either triangular or square wave), the number of individuals with FEV_1 decrement greater than 10% was 2 out of 30 test subjects. This is an extremely small number of individuals from which to conclude that the

response is significant. A direct 95% confidence interval calculation (see Appendix) is that the proportion in the general population who would show a 10% FEV_1 decrement at an ozone level of 0.06 ppm is between 0.8% and 22%. This is quite different from the impression created by Fig. 5.2(a), and indicative of a much higher level of uncertainty.

The staff paper response curve is based on combining data from several ozone levels using a logistic response curve, but this assumes that the response curve fitted to higher ozone levels can be extrapolated downwards to 0.06 ppm. Given the large uncertainty in the probability of response at 0.06 ppm ozone, I do not believe the staff paper's conclusions on this point are justified.

References

Adams, W.C. (2002), Comparison of chamber and face mask 6.6-hour exposures to ozone on pulmonary function and symptoms responses. Inhalation Toxicology **14**, 745–764.

Adams, W.C. (2006), Comparison of chamber 6.6-h exposure to 0.04-0.08 ozone via square wave and triangular profiles on pulmonary responses. Inhalation Toxicology **18**, 127–136.

First exposure pattern	Second exposure pattern					
	.04 triang.	.04 square	.06 triang.	.06 square	.08 square	.08 triang.
Filter	0.002	0.012	0.028**	0.029***	0.061***	0.070***
.04 triangular		0.011	0.026*	0.027***	0.059***	0.068***
.04 square			0.015	0.016	0.048**	0.058***
.06 triangular				0.001	0.033*	0.042***
.06 square					0.032	0.041*
.08 square						0.009

Tabulated is the mean decrement in FEV1 when moving from the first exposure pattern to the second exposure pattern.

* indicates a statistically significant effect at $\alpha = .05$.
** indicates a statistically significant effect at $\alpha = .01$.
*** indicates a statistically significant effect at $\alpha = .001$.

All comparisons based on pairwise t tests. Tests based on exposure .04 (square wave) calculated by combining data from the two Adams papers (2002, 2006): within each sample, the mean FEV1 decrement was calculated between the exposure of interest and the filtered air control; the difference between those means was used to construct the test.

Appendix: Confidence interval calculation

The standard textbook calculation of the confidence interval for a binomial parameter is based on the normal approximation and leads to the interval $\hat{p} \pm 1.96\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$ where $\hat{p} = \frac{X}{n}$ is the sample proportion (n is the sample size, X is the observed number of responses). However in this case with $n = 30,\ X = 2$, this would lead to the absurd confidence interval $(-.023, .156)$. The reason for the failure is that the normal distribution is not applicable for such small n and p.

So we proceed with an exact calculation, as follows. We make use of the following equivalence principle between tests and confidence intervals: a 95% confidence interval for p consists of all values of p_0 for which a test of the hypothesis $H_0 :\ p = p_0$ against the alterative $H_1 :\ p \neq p_0$ would *not* be rejected at significance level .05.

A two-sided, equal tailed test would reject H_0 when $X = x$ if $x \leq x_L$ or $x \geq x_U$ where the lower and upper bounds $x_L,\ x_U$ satisfy

$$\Pr\{X \leq x_L \mid p = p_0\} \ \leq .025, \quad \Pr\{X \geq x_U \mid p = p_0\} \ \leq .025.$$

Therefore, we reject when $x = 2$ if and only if

$$\Pr\{X \leq 2 \mid p = p_0\} \ \leq .025 \text{ or } \Pr\{X \geq 2 \mid p = p_0\} \ \leq .025.$$

But based on $n = 30$, $\Pr\{X \leq 2\} = (1-p)^{30} + 30p(1-p)^{29} + 435p^2(1-p)^{28}$ and $\Pr\{X \geq 2\} = 1-(1-p)^{30} - 30p(1-p)^{29}$. By solving, respectively, $(1-p)^{30} + 30p(1-p)^{29} + 435p^2(1-p)^{28} = .025$ for the upper bound (solution: $p = .22073$) and $(1-p)^{30} + 30p(1-p)^{29} = .975$ for the lower bound (solution: $p = .00818$), we find the confidence interval as stated.