UNITED STATES ENVIRONMENTAL PROTECTION AGENCY
National Center for Environmental Assessment
Research Triangle Park, NC 27711

MEMORANDUM

SUBJECT: The effects of ozone on lung function at 0.06 ppm in healthy adults

FROM: James S. Brown, EPA, NCEA-RTP, Environmental Media Assessment Group (EMAG)

THRU: Mary Ross, EPA, NCEA-RTP, EMAG, Branch Chief
Ila Cote, EPA, NCEA-RTP, Director

TO: Ozone NAAQS Review Docket (OAR-2005-0172)

DATE: June 14, 2007

The purpose of this memo is to address issues that have arisen, including those raised in public comments since issuance of the EPA report, "Review of the National Ambient Air Quality Standards for Ozone: Policy Assessment of Scientific and Technical Information – OAQPS Staff Paper" (hereafter, OAQPS Staff Paper), in January 2007 concerning the controlled human exposure studies published by Dr. Adams which analyzed ozone (O_3)-related lung function decrements. The principal issue raised is the extent to which lung function and respiratory symptom responses observed in healthy adult subjects in Dr. Adams studies at 0.06 ppm O_3 for a 6.6 hour exposure while engaged in moderate exertion were or were not statistically significant.

At the time of the 1996 EPA Ozone Air Quality Criteria Document or O_3 AQCD (U.S. EPA, 1996), statistically significant lung function decrements, viz., forced expiratory volume in one second (FEV_1), and respiratory symptom responses were reported in young healthy adults exposed to 0.08 ppm O_3 for 6.6 hr during quasi continuous exercise (Horstman et al., 1990; McDonnell et al., 1991).[1] In 1996, published studies of ozone exposures at levels below 0.08 ppm were not available. Adams (2002, 2006) provides data for O_3 exposures of 0.04 and 0.06 ppm.[2] Consistent with prior studies, Adams (2002, 2006) reported statistically significant effects of O_3 on FEV_1 and respiratory symptom responses at 0.08 ppm. Below 0.08 ppm O_3, Adams (2006) only reported a significant O_3 effect on a total symptom score (TSS) for the triangular 0.06 ppm O_3 protocol following 5.6 and 6.6 hrs of exposure. There was a tendency, however, as

---

[1] In the 6.6 hr quasi continuous exercise protocol, subjects performed six 50-minute periods of exercise (minute ventilation, ~40 L/min) followed by 10-minutes of each hour while exposed to O_3 or filtered air. Subjects had an additional 35 minutes of rest for a lunch break after the third hour.

[2] The Adams (2006) study also employed the 6.6 hr protocol (see Footnote 1), however, both square-wave (referred to here as "constant") and triangular exposure profiles were utilized. The O_3 concentrations for the 6.6-h exposures were constant at 0.00, 0.06 and 0.08 ppm for square-wave profiles and averaged 0.04, 0.06, and 0.08 ppm for triangular profiles.

illustrated in Figure 1 for the group mean FEV_1 responses during the 0.06 ppm O_3 exposure to diverge from the responses observed at 0.0 and 0.04 ppm O_3.

The Adams (2006) study investigated the effects of 6.6 hour square-wave (0.000, 0.060, and 0.080 ppm O_3) and triangular (averaging 0.040, 0.060, and 0.080 ppm O_3) exposures on lung function and respiratory symptoms during intermittent exercise in 30 healthy young adults. The study design compared FEV_1, TSS, and pain on deep inspiration (PDI) between the six exposure protocols at each of six time points (1, 2, 3, 4.6, 5.6, and 6.6 hours). The author was principally interested in evaluating the pattern of responses at each time interval and, therefore, conducted a two-way analysis of variance with repeated measures. A conservative statistical test, the Scheffé post hoc test, was used by the author to minimize Type I errors (falsely rejecting the null hypothesis of no difference) when performing multiple comparisons.

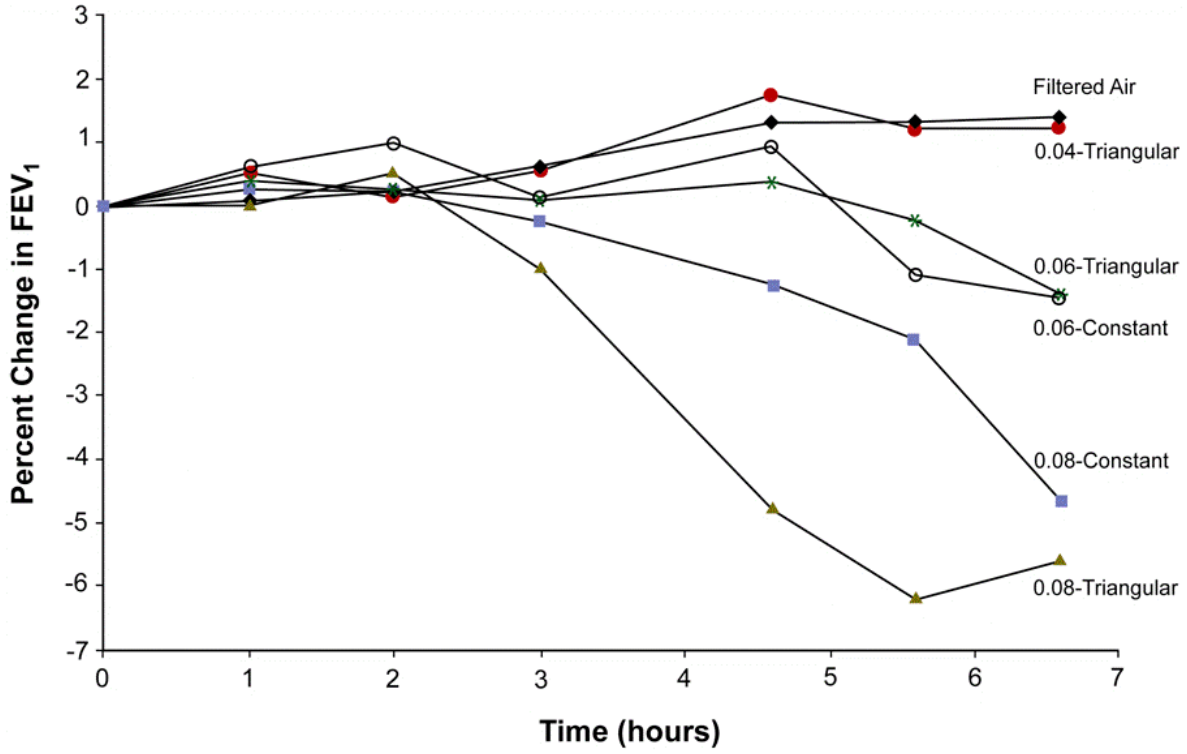

Figure 1. Hour by hour changes in FEV_1 (% change relative to preexposure) adapted from Adams (2006). Data are group mean (error bars were not provided in the published paper) responses of 30 healthy adults exposed to O_3 for 6.6 hours during quasi continuous exercise. The O_3 concentrations were either held constant for the entire 6.6 hour exposure or gradually increased to the lunch hour and then decreased to give a triangular exposure profile of an average concentration noted in the figure.

At 6.6 hours, FEV_1 responses from both square-wave and triangular 0.080 ppm O_3 exposures were found to be statistically significantly different from the responses observed for 0.000, 0.040, and 0.060 ppm O_3 exposures. Compared to preexposure, FEV_1 responses for the triangular 0.080 ppm protocol were statistically significantly different at 4.6 hours, while the square-wave 0.080 ppm responses were statistically significantly different at 6.6 hours. The author reported that hourly changes in FEV_1 responses in both square-wave and triangular 0.060 ppm O_3 exposures did not differ significantly from each other, nor were they statistically significantly different from the filtered air (0.000 ppm O_3) responses at 6.6 hours of exposure. For FEV_1 responses, triangular exposure to 0.04 ppm O_3 also was not statistically significantly different from the filtered air response throughout the protocols. Furthermore, Adams also reported that TSS values during square-wave and triangular 0.080 ppm O_3 exposure reached statistical significance relative to preexposure at 5.6 and 4.6 hours, respectively. The triangular 0.060 ppm O_3 exposure reached statistical significance by 5.6 hours, whereas the square-wave 0.060 ppm exposure did not approach statistical significance by 6.6 hours. The author stated that PDI values followed a similar pattern to the TSS.

In the OAQPS Staff Paper, EPA examined pair-wise comparisons of the group mean FEV_1 responses in Figure 1 of Adams (2006). The visual comparison suggested that responses during the 0.060 ppm O_3 exposures appear to diverge from responses for filtered-air and 0.040 ppm O_3 (EPA, 2006, p. 8-42). We were concerned that, in addition to reducing the probability of Type I error (false positive), the correction for the multiple comparisons by Adams (2006) may have also increased Type II error (false negative) for the simple evaluation of pre- to postexposure effects of O_3 versus filtered air on FEV_1, as has been commonly assessed by others (e.g., Horstman et al., 1990; McDonnell et al., 1991). As discussed in more detail in Section 3.3.1.1.1 of the OAQPS Staff Paper, the staff's cursory evaluation of pre- to postexposure effects found there was a lack of an overlap in the range of responses (i.e., the means± Standard Error (SE)) at 0.060 ppm O_3 versus filtered air at 6.6 hr, and this is suggestive of a statistically significant effect on FEV_1.

Subsequent to release of the OAQPS Staff Paper in late January 2007, public comments were submitted by Dr. Richard Smith and he summarized these comments in a presentation at the March 5, 2007 CASAC teleconference to discuss the Staff Paper.[3] Dr. Smith noted, "…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…" He reported the results of comparisons using a range of different statistical techniques to demonstrate the sensitivity of the results to the underlying assumptions. Consistent with common practice for comparing pre- and postexposure responses to test for whether or not an O_3-related effect is significant, Dr. Smith used a conventional paired *t* test. Dr. Smith's analysis shows that the small (< 3 percent)

[3] Public comment submitted by Dr. Richard Smith on March 4, 2007 for presentation at the March 5, 2007 CASAC teleconference to discuss the Staff Paper released on January 31, 2007 (docket number: EPA–HQ-OAR-2005-0172-0080). Dr. Smith Public comment docket number: EPA–HQ-OAR-2005-0172-0080.

group mean FEV_1 decrement following the 6.6-hour exposure at 0.060 ppm is statistically significantly ($p \leq 0.01$) different from filtered air responses using this approach.[4]

In this memo we address and replicate the analysis submitted by Dr. Smith and address the issues raised in his public comments concerning the statistical significance of 0.06 ppm O_3 exposure on FEV_1 in the Adams (2006) publication. Studies conducted by the U.S. EPA in Chapel Hill, NC have commonly utilized a paired *t* test to assess the statistical significance ($p \leq 0.05$) of pre- to postexposure changes in FEV_1 between an air and an O_3 exposure (e.g., Horstman et al., 1990; McDonnell et al., 1991). To assess the "true" effect of O_3, the air exposure controls for a variety of factors such as exercise, intrasubject variability in baseline conditions, and effects of the laboratory exposure setting itself. Such an approach is standard for both short term (1-2 hr) exposures and prolonged (6.6-8 hr) exposures assessing the effects of O_3 on lung function as well as for testing differences in responses between healthy and diseased individuals such as asthmatics. Adams (2002, 2006) utilized a more conservative technique that was intended to address the comparisons related to the time course of the responses which was the primary research question in his study. The goal here is not to critique the statistical approaches of any study, but rather: 1) to note differences in the statistical methods between studies and 2) to analyze FEV_1 responses to low O_3 exposure concentrations from the Adams' studies in the same manner as the studies conducted by the U.S. EPA in Chapel Hill, NC.

As already stated, in contrast to simply testing pre- to postexposure effects on FEV_1, Adams (2006) analyzed for statistical significance ($p < 0.05$) using a two-way analysis of variance (ANOVA) with repeated measures, which tested for gas concentration (including square-wave or triangular exposure mode) effects and exposure time effects. The Scheffé post hoc test (Kleinbaum et al., 1988) was applied to determine which particular mean values were significantly different from each other. Adams (2006) utilized this statistical approach to correct for multiple comparisons between O_3 exposure concentrations (0, 0.04, 0.06, and 0.08 ppm), exposure profile (square-wave and triangular), and exposure time (0, 1, 2, 3, 4.6, 5.6, and 6.6 hr). Corrected for the multiple comparisons, Adams (2006) reported significant reductions in FEV_1 (relative to preexposure) and TSS for the 0.08 ppm O_3 protocols at 4.6 h and thereafter. TSS were also significantly increased relative to baseline for the triangular exposure to 0.06 ppm O_3 and at 5.6 and 6.6 hrs.

In the Adams (2006) study, assuming the FEV_1 responses at each time point were adjusted to preexposure values (t=0), it is important to note that the post hoc test corrected for 90 comparisons (15 protocol comparisons at each of 6 time points). Correcting for multiple comparisons avoids rejecting the null hypothesis when it is true. Unfortunately, reducing the

[4] Dr. Smith expressed the view that the appropriate comparison should be between the 0.060 and 0.040 ppm levels, where 0.040 ppm was selected to represent a background level, rather than filtered air and he found more mixed results in terms of statistical significance using this comparison. As discussed below, we and most authors of the controlled human exposure studies believe that the appropriate approach for testing for an O_3-related response is to compare with filtered air to correct for the effect of exertion in clean air. Additionally, as discussed in the O_3 AQCD (EPA, 2006, AX3-131) and in Chapter 2 of the OAQPS Staff Paper, the scientific evidence supports estimates of policy-relevant background that are in the 0.015 to 0.035 ppm range in the afternoon during the O_3 warm season, rather than the 0.040 ppm level cited by Dr. Smith.

Type I error (falsely rejecting the null) increases the Type II error (falsely accepting the null). For example, applying a simple *Bonferroni correction* to the Adams (2006) scenario, the critical p-value for rejecting the null hypothesis of no change in FEV_1 would be 0.05/90 or 0.000556. By contrast, a critical p-value might more appropriately be 0.05/5 or 0.01 for assessing pre- to postexposure changes in FEV_1 between an air and an O_3 exposure in the Adams (2006) study. We conclude that, although appropriate for the design and intent of the Adams' studies, the multiple comparison correction is overly conservative (increased Type II error and decreased power) for the evaluation of pre- to postexposure changes in FEV_1 between an air and an O_3 exposure and we adopted the standard approach used by other researchers (e.g., Hazucha et al., 1992; Horstman et al., 1995; McDonnell et al., 1991).

Dr. Adams submitted public comments and summarized these comments during the CASAC O_3 Panel's March 5, 2007 teleconference in which he questioned the approach used in the OAQPS Staff Paper involving the comparison of standard errors to assess whether the lung function and TSS responses observed in his 2006 article were likely statistically significantly different for the 0.060 ppm scenario compared to filtered air. Dr. Adams expressed the view that the standard deviation (SD) reported in Adams (2006) for both lung function responses and TSS should be used instead of SE. On the March 5, 2007 teleconference, members of the CASAC O_3 Panel noted the very conservative nature of the statistical test used by Adams to evaluate the research questions posed by the author. These same CASAC Panel members also supported the approach adopted in the OAQPS Staff Paper to evaluate the statistical significance of O_3-related lung function responses associated with pre- versus postexposure responses. The CASAC Panel members also supported the use of the paired *t* test approach as the preferred method for analyzing the pre- minus postexposure lung function responses.

EPA staff's analysis is summarized in Attachment 1 and shows results where the paired *t* test is used to compare whether group mean FEV_1 responses associated with 0.06 ppm exposure (both square wave and triangular exposures) are statistically significantly different compared to group mean responses associated with filtered air. The results of EPA's analysis for these comparisons confirm the results presented by Dr. Richard Smith in his public comments. For the comparison of the 0.06 ppm square wave exposure versus filtered air, the difference in means is statistically significant at the 99.9% confidence level (i.e., p-value = 0.001) for the two-tailed test. Similarly, for the comparison of the 0.06 ppm triangular exposure versus filtered air, the difference in means is statistically significant at the 99% confidence level (i.e., p-value < 0.01). Individual data used in this analysis were obtained from Dr. Adams for the purposes of conducting the lung function health risk assessment. We conclude that the pre- to postexposure analysis conducted here shows that exposure to 0.06 ppm O_3 also causes a relatively small but statistically significant decrease (post- minus preexposure) in group mean FEV_1 responses compared to filtered air. Furthermore, Figure 2 illustrates the effects of 0.06 ppm O_3 exposure are consistent with the trend in responses observed for exposures to 0.04 and 0.08 ppm O_3.

As illustrated in Figure 2, the average FEV_1 response to 0.06 ppm O_3 exposure is relatively small, but is important as this is an average response in young healthy adults. As observed in Attachment 1, there is considerable variability in responses between similarly exposed individuals, such that some experience distinctly larger effects even when small group mean

responses are observed. Many factors such as age, gender, disease, nutritional status, smoking, and genetic variability may contribute to the differential health effects of O_3 exposure. More detail on intersubject variability is available in Sections 6.3-6.5 and 8.7 of the O_3 AQCD (U.S. EPA, 2006). Larger decrements in FEV_1 than described here might be expected in more susceptible populations.

In summary, exposure to 0.06 ppm O_3 causes a relatively small but statistically significant decrease (post- minus preexposure) in group mean FEV_1 responses in young healthy adults compared to filtered air responses. Some healthy individuals experience moderate (>10%) decrements in FEV_1 when exposed to 0.06 ppm O_3 relative to filtered air (see Attachment 1). As noted by Adams (2006), TSS are also increased relative to baseline by 5.6 hrs of exposure to 0.06 ppm O_3. Based on the current body of literature, it is reasonable to expect susceptible populations, such as age-matched asthmatics, to experience at least equivalent or greater decrements in FEV_1. It would further be expected (EPA, 2006, p. 8-68), that asthmatics experiencing moderate responses to 0.06 ppm O_3 exposure would limit their activity and increase their frequency of medication usage.

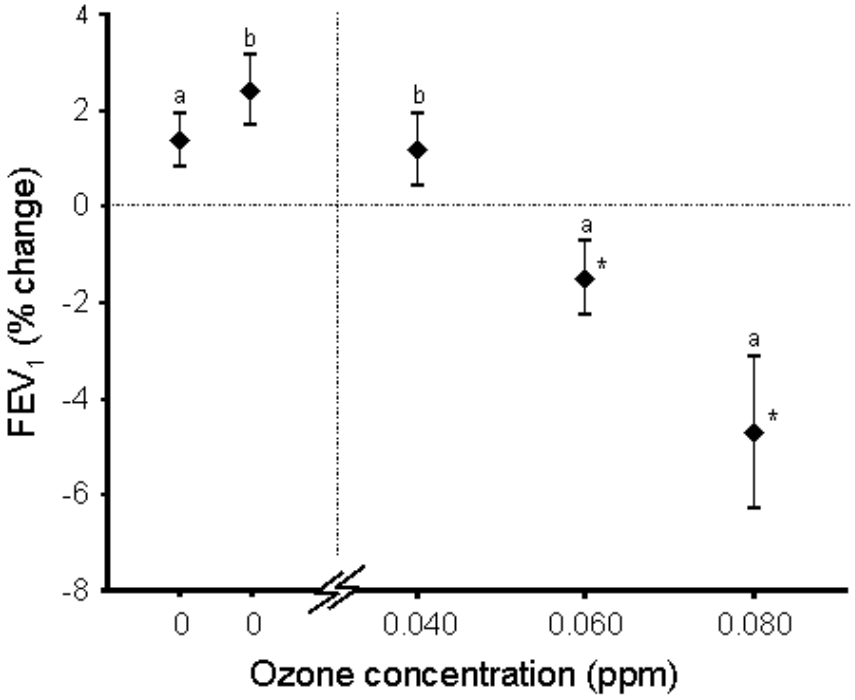

Figure 2. Effects of ozone on FEV_1 in healthy young adults exposed for 6.6 h during quasi continuous exercise to a constant (square-wave) O_3 concentration. Data are from a) Adams (2006) and b) Adams (2002). *Significantly different from responses to air exposure ($p \leq 0.001$, two-tail paired *t* test).

References:

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

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 Toxicol. 18: 127-136.

Hazucha, M. J.; Folinsbee, L. J.; Seal, E., Jr. (1992) Effects of steady-state and variable ozone concentration profiles on pulmonary function. Am. Rev. Respir. Dis. 146: 1487-1493.

Horstman, D. H.; Folinsbee, L. J.; Ives, P. J.; Abdul-Salaam, S.; McDonnell, W. F. (1990) Ozone concentration and pulmonary response relationships for 6.6-hour exposures with five hours of moderate exercise to 0.08, 0.10, and 0.12 ppm. Am. Rev. Respir. Dis. 142: 1158-1163.

Horstman, D. H.; Ball, B. A.; Brown, J.; Gerrity, T.; Folinsbee, L. J. (1995) Comparison of pulmonary responses of asthmatic and nonasthmatic subjects performing light exercise while exposed to a low level of ozone. Toxicol. Ind. Health 11: 369-385.

Kleinbaum, D. G., Kupper, L. L., and Muller, K. W. 1988. Applied Regression Analysis and Other Multivariable Methods. pp. 368-372. Boston, MA: PWS-Kent.

McDonnell, W. F.; Kehrl, H. R.; Abdul-Salaam, S.; Ives, P. J.; Folinsbee, L. J.; Devlin, R. B.; O'Neil, J. J.; Horstman, D. H. (1991) Respiratory response of humans exposed to low levels of ozone for 6.6 hours. Arch. Environ. Health 46: 145-150.

U.S. Environmental Protection Agency. (1996) Air quality criteria for ozone and related photochemical oxidants. Research Triangle Park, NC: Office of Research and Development; report nos. EPA/600/AP-93/004aF-cF. 3v. Available: http://cfpub2.epa.gov/ncea.

U.S. Environmental Protection Agency. (2006) Air quality criteria for ozone and related photochemical oxidants. Research Triangle Park, NC: Office of Research and Development; report no. EPA/600/R-05/004aF. Available: http://cfpub2.epa.gov/ncea.

U.S. Environmental Protection Agency. (2007). Review of the National Ambient Air Quality Standards for Ozone: Policy Assessment of Scientific and Technical Information, OAQPS Staff Paper; report no. EPA/452/R-07/03. Available: http://www.epa.gov/ttn/naaqs/standards/ozone/s_o3_cr_sp.html

Attachment 1

Adams (2006) -- Fraction change in FEV_1 following 0.06 ppm ozone exposure and filter air

Subj	Filtered Air (FA)	0.06 ppm O_3 square wave	Difference between 0.06 ppm and FA	0.06 ppm O_3 triangular	Difference between 0.06 ppm and FA
1	0.0146	0.0087	-0.0059	0.0028	-0.0117
2	-0.0020	-0.0301	-0.0281	0.0207	0.0226
3	0.0029	-0.0593	-0.0623	-0.0334	-0.0364
4	0.0533	0.0352	-0.0181	-0.0352	-0.0885
5	0.0862	0.0375	-0.0487	0.0974	0.0112
6	0.0294	0.0389	0.0095	-0.0029	-0.0323
7	0.0021	-0.0021	-0.0042	0.0191	0.0169
8	0.0028	-0.0343	-0.0371	0.0227	0.0199
9	0.0312	-0.0391	-0.0703	0.0234	-0.0078
10	0.0632	-0.0643	-0.1276	-0.1198	-0.1830
11	0.0053	-0.0053	-0.0107	0.0000	-0.0053
12	-0.0476	0.0266	0.0742	-0.0812	-0.0335
13	0.0171	0.0287	0.0116	0.0021	-0.0150
14	0.0042	-0.0287	-0.0329	0.0000	-0.0042
15	0.0116	0.0057	-0.0058	0.0472	0.0357
16	0.0554	0.0174	-0.0380	-0.0248	-0.0802
17	0.0062	0.0105	0.0043	-0.0021	-0.0083
18	0.0435	-0.0188	-0.0623	0.0377	-0.0058
19	0.0317	-0.0198	-0.0515	-0.0493	-0.0810
20	-0.0202	-0.0503	-0.0300	-0.0104	0.0098
21	0.0219	-0.0366	-0.0585	0.0264	0.0045
22	0.0219	0.0513	0.0293	-0.0296	-0.0515
23	-0.0021	-0.1473	-0.1452	-0.0783	-0.0761
24	0.0058	-0.0059	-0.0117	0.0182	0.0123
25	0.0027	-0.0054	-0.0081	-0.0583	-0.0611
26	0.0414	-0.0110	-0.0524	0.0385	-0.0029
27	-0.0179	-0.0202	-0.0023	0.0051	0.0231
28	-0.0019	0.0040	0.0060	0.0344	0.0364
29	-0.0087	-0.0421	-0.0334	-0.0860	-0.0772
30	-0.0502	-0.0962	-0.0460	-0.2138	-0.1637
Mean	0.0135	-0.0151	**-0.0285**	-0.0143	**-0.0278**
StDev	0.0298	0.0424	0.0428	0.0595	0.0541
StErr	0.0054	0.0077		0.0109	

t-Test: Paired Two Sample for Means
Square-wave profile versus filtered air

	Variable 1	*Variable 2*
Mean	0.013461	-0.015072
Variance	0.000886	0.001796
Observations	30	30
Pearson Correlation	0.338561	
Hypothesized Mean Difference	0	
df	29	
t Stat	3.654981	
P(T<=t) one-tail	0.000506	
t Critical one-tail	1.699127	
P(T<=t) two-tail	**0.001012**	
t Critical two-tail	2.04523	

t-Test: Paired Two Sample for Means
Triangular profile versus filtered air

	Variable 1	*Variable 2*
Mean	0.013461	-0.014309
Variance	0.000886	0.003536
Observations	30	30
Pearson Correlation	0.422164	
Hypothesized Mean Difference	0	
df	29	
t Stat	2.81103	
P(T<=t) one-tail	0.00438	
t Critical one-tail	1.699127	
P(T<=t) two-tail	**0.00876**	
t Critical two-tail	2.04523	