Paper PO-139

Evaluating Additivity of Health Effects of Exposure to Multiple Air Pollutants Given Only Summary Data

Laura Datko Williams, ORISE/U.S. EPA; Elizabeth Oesterling Owens, U.S. EPA;
Jean-Jacques Dubois, Critical Path Services, LLC

ABSTRACT

A research team is interested in determining whether or not health effects of exposure to a mixture of air pollutants is additive, using summary statistics found in published toxicology studies. Additivity is defined as no significant difference between the effects of exposure to the mixture and the sum of the effects of exposure to each individual component of that mixture. The studies of interest to the research team typically did not test for this difference, even though the study design often made it possible. Many, however, did provide three summary statistics sufficient to reproduce the test exactly (number of subjects [n], mean response, and standard deviation). Using SAS®, we can perform the analysis of interest given those summary statistics. First, SAS/STAT® can be used to generate datasets standardized to those three statistics for each study. Next, using the appropriate ESTIMATE in PROC GLM, the effect of the mixture of pollutants can be tested against the sum of the effects of each component of the mixture. In order to compare results between studies of different toxicological endpoints, a relative difference between the mixture and the sum can be calculated. Next, the IML procedure can be used to calculate confidence intervals. Finally, a convenient way to display the results is with a forest plot, which can be created using PROC SGPLOT (SAS® ODS Graphics). Study details can be added to the plot as data points. Here we present the method we developed to test for the effect of interest to the research team: is the effect of the mixture equal to the sum of individual component effects? This method allowed us to obtain the exact test results that would have been obtained for the effect of interest using the original full data, if the original authors had conducted that test.

INTRODUCTION

The U.S. Environmental Protection Agency (EPA) regulates several common air pollutants, or criteria air pollutants, through the implementation of the National Ambient Air Quality Standards (NAAQS). The primary purpose of the NAAQS is to protect public health, accounting for sensitive sub-populations, with an adequate margin of safety. The Clean Air Act requires periodic updating of the NAAQS, based on any new scientific evidence supporting the levels of the standards. Several recent reports and publications (Johns et al., 2012; Hidy et al., 2011; Mauderly et al., 2010; NRC, 2004), have suggested that to support air quality regulations, the effects of the criteria pollutants should be evaluated as a mixture, rather than individually. A research team at the U.S. EPA is interested in evaluating published data from peer-reviewed studies of the health effects of mixtures of criteria air pollutants. Specifically, the interest is in testing the research hypothesis that the sum of effects due to each individual component of the mixture is not equal to the effects of exposure to the mixture as a whole.

Generally, the default assumption in toxicology research is that the mixture of pollutants has an additive effect on the health outcome of interest. In other words, the sum of responses to each individual component of the mixture is generally believed to be equal to the response from exposure to the mixture as a whole. However, this is rarely tested, even though the design of many experimental studies would allow it to be. We gathered a collection of such studies published in peer-reviewed journals, in which the data was almost always provided only through a minimal set of summary statistics. We present a method for using reported summary statistics (i.e. number of experimental units, mean, and standard deviation or standard error) to test for the effect of interest (does sum = mixture?), obtaining the exact results that would have been obtained from the primary data by the original authors of the publications. We also include some data from our research project to illustrate this method. We are using this method to provide a mathematical foundation to claims of additive effects of mixtures of air pollutants and to synthesize the results of many studies.

GENERATING DATASETS FOR ANALYSIS

As pointed out by Larson (1992), all data sets with the same number of observations, mean and standard deviation gives identical analysis of variance (ANOVA) results. Several methods have been proposed to generate data sets from those sufficient three summary statistics. Lehman (1993), in response to Larson's remarks, suggested that those synthetic data sets would have a more pleasant appearance if they were also random and normally distributed, although these two properties are not required in order to obtain the same ANOVA results. Following these methods, we generated data sets from the summary data extracted from the published toxicology studies. The studies had 4 experimental groups: null (control) exposure, exposure to pollutant A (A_alone), exposure to pollutant B (B_alone), and exposure to the mixture of A and B (mixture). The following table shows an example of the source data file.

Study_ID	Exposure	N	Mean	StDev
1201_a	control	4	101	5
1201_a	A_alone	4	221	70
1201_a	B_alone	4	283	62
1201_a	mixture	4	543	55
1252_g	control	4	2.671	0.329
1252_g	A_alone	4	2.928	0.395
1252_g	B_alone	4	5.862	0.197
1252_g	mixture	4	6.207	0.307
1252_h	control	4	0.373	0.034
1252_h	A_alone	4	0.387	0.0227
1252_h	B_alone	4	0.435	0.017
1252_h	mixture	4	0.498	0.017
15440_d	control	10	6.39	0.77
15440_d	A_alone	8	6	1.77
15440_d	B_alone	8	7.97	0.63
15440_d	mixture	4	9.8	1.06

Table 1: Example Source Data Set

The following SAS® macro generates a data set from the three sufficient statistics, one data set for each line of the source data. Random observations are generated and they are then normalized and standardized to the provided mean and standard deviation. Note the following macro variables: FILE is the source data set, VAR1 is the study identifier, VAR2 is the exposure group, VAR3 is “n” or number of subjects, VAR4 is the mean, VAR5 is the standard deviation.

```
%MACRO MAKEDATA(FILE, VAR1, VAR2, VAR3, VAR4, VAR5);
	options nonotes;

	data _NULL_;
		if 0 then set &FILE NOBS=nobs;
		call symput('OBSCOUNT', nobs);
		stop;
	run;

	%do I=1 %to &OBSCOUNT;
		data _NULL_;
		set &FILE (FIRSTOBS=&I);
			call symput('study', &VAR1);
			call symput('cell', &VAR2);
			call symput('n', &VAR3);
			call symput('mean', &VAR4);
			call symput('stdev', &VAR5);
			stop;
		run;

		data A;
			do i=1 to &n;
				y=rand('normal', 0, 1);
				study="&study";
				cell="&cell";
				output;
			end;
			drop i;
		run;

		proc standard data=A mean=&mean std=&stdev out=A;
		run;
```

```
            proc append base=SIMULATE data=A;
            run;
      %END;
%MEND MAKEDATA;

%MAKEDATA(SOURCE, study, cell, n ,mean, stdev)
```

First the macro determines how many rows are in the source data file. Then, the do-group loops through from the first source observation to the last, doing the following:

1. Creating the macro variables study, cell (i.e. treatment group), n, mean and stdev
2. Taking the current row of the source data set and generating a new data set (A) with *n* observations from a normal distribution.
3. Standardizing A to the mean and standard deviation from the source data set.
4. Appending A to a new data set called SIMULATE

When the macro has finished, the dataset SIMULATE contains the data generated from each row of the source data file. Therefore if the first row of Table 1 is fed into the macro, SIMULATE will have 4 corresponding observations, with a mean of 101 and a standard deviation of 5.

TESTING FOR ADDITIVITY

All the studies that were to be evaluated for additivity had a completely randomized design with a factorial arrangement of two pollutants (pollutant A and pollutant B) at two levels each (present or absent). The subjects were randomly assigned to the following four treatment groups: control exposure (A0B0), exposure to pollutant A (A1B0), exposure to pollutant B (A0B1), or exposure to a mixture of pollutant A and pollutant B (A1B1). Analysis of variance was performed, at $\alpha = 0.05$, on the SIMULATE data using PROC GLM. The null and alternative hypotheses are as follows:

$$H_0: (\mu_A + \mu_B) = \mu_M$$
$$H_a: (\mu_A + \mu_B) \neq \mu_M$$

Where μ_A represents the mean response from exposure to pollutant A (A1B0 – A0B0), μ_B represents the mean response from exposure to pollutant B (A0B1 – A0B0), and μ_M represents the mean response from exposure to the mixture of A and B (A1B1 – A0B0). The effect due to control (null) exposure was subtracted from each of these values. The following code presents the use of PROC GLM for a 2 X 2 factorial model.

```
ods output  means=CELLMEANS estimates=RESULTS overallanova=ROOTS;
ods html close;
ods listing close;

proc GLM data=SIMULATE  plots=none; by study;
      class A B ;
      model Y = A|B / CLPARM;
      means A|B;
      estimate 'A1B0-A0B0' A -1 1 A*B -1 0 1 0;
      estimate 'A0B1-A0B0' B -1 1 A*B -1 1 0 0;
      estimate '(A1B0-A0B0) + (A0B1-A0B0)' A -1 1 B -1 1 A*B -2 1 1 0;
      estimate 'A1B1-A0B0' A -1 1 B -1 1 A*B -1 0 0 1;
      estimate '(A1B1-A0B0)-(A0B1-A0B0+A1B0-A0B0)' A*B 1 -1 -1 1;
run;
quit;
ods output close;
ods listing;
```

From the PROC GLM output, several important results can be determined for each experiment in the source data:

- The first two ESTIMATE statements provide tests of the effect of exposure to each pollutant by itself, the third ESTIMATE provides a test of the effect of the sum of exposures to each pollutant, and the fourth ESTIMATE provides a test of the effect of exposure to the mixture. Each of these effects is tested as a difference from the control group.
- The fifth ESTIMATE statement provides a test of whether the effect of the sum of exposures to individual pollutants is different from the effect of exposure to the mixture. If the null hypothesis was not rejected, we concluded the effects of the mixture were not significantly different from additive.

- If none of the exposure groups (A alone, B alone, mixture) had an effect different from the effect of control (null) exposure, that experiment was flagged. These experiments were classified as having no effects for any exposure.

Next, we wanted to be able to compare the direction and magnitude of the mixture effects, relative to additive, across the different studies. The studies measured different types of effects at various doses of the mixtures, therefore we decided to calculate a unitless ratio to represent the relative difference from additive:

$$\text{Quotient} = \frac{\text{effects of mixture - sum of individual effects}}{\text{sum of individual effects}}$$

We utilized the methods of Dilba et al. (2006) to calculate the ratios and 95% confidence limits for them, based on Fieller's theorem. The datasets CELLMEANS, RESULTS, and ROOTS, obtained from PROC GLM using ODS OUTPUT, are first used to prepare the data necessary to compute the confidence intervals. Some of the data could alternatively be taken from the SOURCE dataset.

```
data TEMP1;
      set CELLMEANS;
      where Effect="A_B";
      cell=A||B;
      Ninv=1/N;
      keep study Ninv mean_y cell;
run;

proc transpose data=TEMP1 out=TEMP2 prefix=Ybar;
      var mean_y;
      id cell;
      by study;
run;

proc transpose data=TEMP1 out=TEMP3 prefix=n;
      var Ninv;
      id cell;
      by study;
run;

data TEMP4;
      merge TEMP2 TEMP3 ROOTS (where=(Source="Error"));
      by study;
      drop _name_ _label_ source dependent SS Fvalue probF;
run;
quit;
```

Next, the SAS® macro FIELLER uses PROC IML to calculate confidence intervals for the relevant ratio. Note the following macro variables: FILE is the prepared dataset from the previous steps (TEMP4), VAR1 is the study identifier, VAR2 through VAR5 are the means for control, A, B, and mixture, VAR6 through VAR9 are the n's for control, A, B, and mixture, VAR10 is the degrees of freedom, and VAR11 is the overall ANOVA MSE.

```
%MACRO
FIELLER(FILE,VAR1,VAR2,VAR3,VAR4,VAR5,VAR6,VAR7,VAR8,VAR9,VAR10,VAR11);
      options nonotes;
            data _NULL_;
                  if 0 then set &FILE NOBS=nobs;
                  call symput('OBSCOUNT',nobs);
            stop;
            run;

      %do I=1 %to &OBSCOUNT;
            data _NULL_;
                  set &FILE (FIRSTOBS=&I);
                  call symput('study',&VAR1);
```

```
                call symput('Ybar11',&VAR2);
                call symput('Ybar12',&VAR3);
                call symput('Ybar21',&VAR4);
                call symput('Ybar22',&VAR5);
                call symput('ninv11',&VAR6);
                call symput('ninv12',&VAR7);
                call symput('ninv21',&VAR8);
                call symput('ninv22',&VAR9);
                call symput('DF',&VAR10);
                call symput('MSE',&VAR11);
          stop;
          run;

          proc iml;
                Ybar = {&Ybar11,&Ybar12,&Ybar21,&Ybar22};
                M = {&ninv11 0 0 0, 0 &ninv12 0 0, 0 0 &ninv21 0, 0 0 0
                &ninv22};
                c = {1,-1,-1,1};
                d = {-2,1,1,0};
                t = tinv(1-0.05/2,&df);
                MSE = &MSE;
                A = (d`*Ybar)**2 - (t**2 * MSE * d` * M * d);
                B = -2*((c`*Ybar)*(d`*Ybar) - (t**2 * MSE * c` * M * d));
                C = (c`*Ybar)**2 - (t**2 * MSE * c` * M * c);
                discr=((B**2)-4*A*C);
                      if (A<=0) then do;
                            lowerbound=.;
                            upperbound=.;
                            stop;
                      end;
                      else do;
                            lowerbound= (-B - sqrt(discr))/(2*A);
                            upperbound= (-B + sqrt(discr))/(2*A);
                            stop;
                      end;
                study= {"&study"};
                bounds= lowerbound||upperbound;
                cname = {"LCL" "UCL"};
                create CI from bounds [colname=cname];
                append from bounds;
                sname={"study"};
                create STUDY from study [colname= sname];
                append from study;
          run;

          data CI; merge STUDY CI;
          run;

          proc append base=ALLCIs data=CI;
          run;
          quit;
     %END;
%MEND FIELLER;

%FIELLER(TEMP4,study,YbarA1B1,YbarA1B2,YbarA2B1,YbarA2B2,nA1B1,nA1B2,nA2B1,nA
2B2,DF,MS)
```

Similar to the MAKEDATA macro, the FIELLER macro first determines the number of observations in the input file, then loops through the data creating macro variables based on the input variables. Next, PROC IML is used to compute the confidence intervals for the ratio of interest.

- Ybar is the column vector of four cell means (Ybar11, Ybar12, Ybar21, Ybar22)
- M is the diagonal matrix of 1/n for the four cells (n11, n12, n21, n22)
- c is the column vector of contrast coefficients for cell means in the numerator of the ratio of interest
- d is the column vector of contrast coefficients for cell means in the denominator of the ratio of interest
- t is the 1-α/2 quantile of a t distribution with degrees of freedom equal to denominator degrees of freedom in the overall ANOVA
- MSE is the overall ANOVA MSE
- A, B, and C are described in Dilba et al. 2006
    - If $A \leq 0$: No solution in this case. Almost always, this occurs when the denominator is non-significant (In this case, the denominator is the sum of effects).
    - Otherwise: The confidence bounds are the two solutions of $Ay^2 + By + C = 0$

Finally, the results are output to a data set and matched to the corresponding Study ID. This ends the IML procedure. The confidence intervals from all studies are compiled into a dataset called ALLCIs.

After running the macro, the results are merged with the output from the PROC GLM and identifying information from the original source data into a final results file. This final file includes, among other parameters, the p-values from the ANOVA, a classification of the results based on the p-value and effect estimates, the quotient, and confidence intervals.

DISPLAYING THE RESULTS

For the intended audience, the most informative way to display the results is in a forest plot. PROC SGPLOT was used to plot multiple scatter plots in a single plot area, using more than one x-axis. The following code can be used to create a forest plot. The variables here correspond with those in Tables 2 and 3. Note that the xaxis and x2axis offsets will need to be adjusted to fit the data on the corresponding plot.

```
proc sgplot data=forest noautolegend;
      scatter y=study x=quotient / group=additivity
                                   xerrorupper=UCL
                                   xerrorlower=LCL
                                   name="plot1";
      scatter y=study x=Reference / markerchar=ref x2axis;
      scatter y=study x=Species / markerchar=spec x2axis;
      scatter y=study x=Endpoint_Type / markerchar=endpt x2axis;
      scatter y=study x=NO2 / markerchar=Nconc x2axis;
      scatter y=study x=O3 / markerchar=Oconc x2axis;
      scatter y=study x=Duration / markerchar=time x2axis;
      xaxis offsetmin=0.75 offsetmax=0 min=-15 max=15
            minor display=(nolabel);
      x2axis offsetmin=0.07 offsetmax=0.35 display=(noticks nolabel noline);
      yaxis reverse display=(noticks nolabel novalues noline);
      refline 0 / axis=x lineattrs=(pattern=shortdash)
                  transparency=0.5;
      keylegend "plot1" / position=topright down=2;
run;
```

Table 2 and 3 show an example of the data (DATA = FOREST) required to create the forest plot. Table 2 and 3 are two parts of a single dataset.

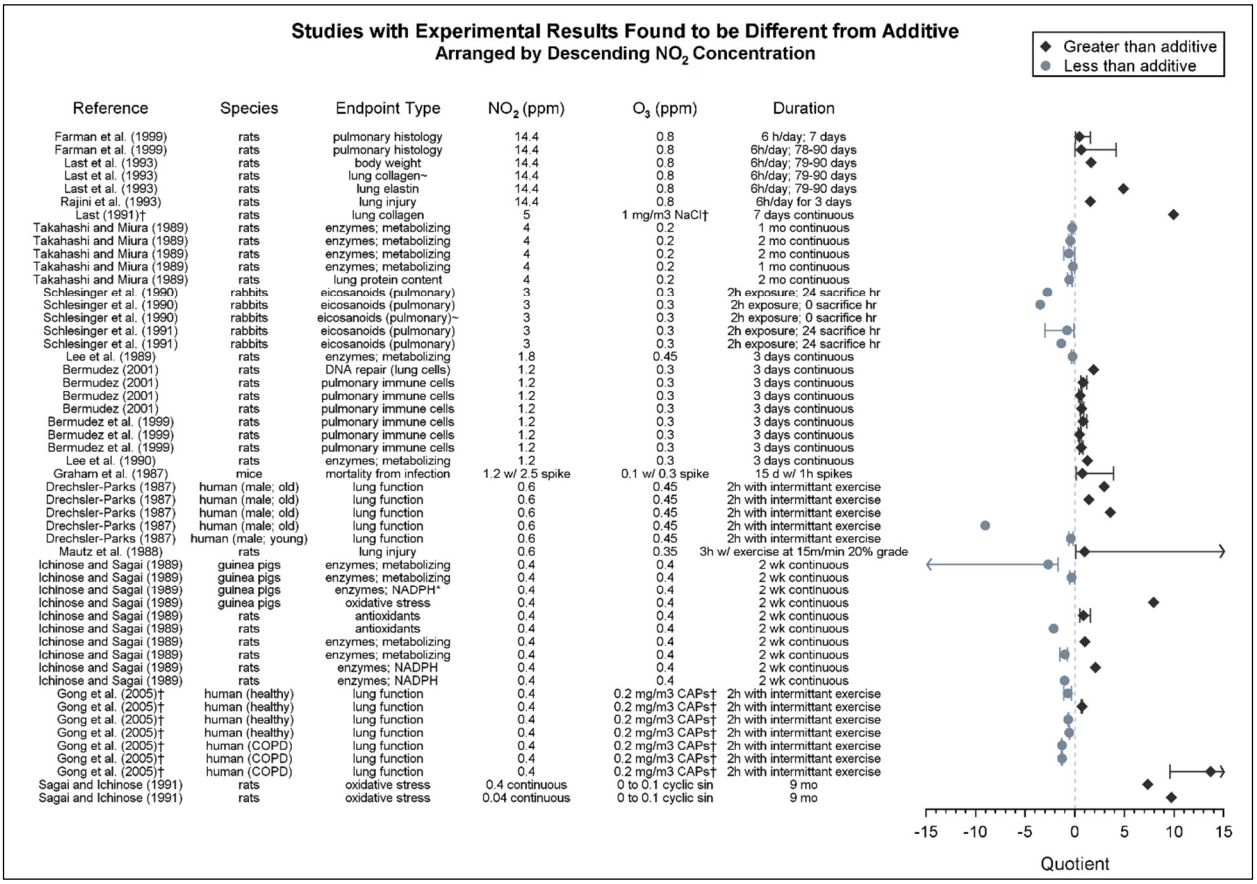
study	ref	spec	endpt	Nconc	Oconc	Time
1201_a	Farman et al. (1999)	rats	pulmonary histology	14.4	0.8	6 h/day; 7 days
1201_b	Farman et al. (1999)	rats	pulmonary histology	14.4	0.8	6h/day; 78-90 days
43705_b	Last et al. (1993)	rats	body weight	14.4	0.8	6h/day; 79-90 days
43705_i	Last et al. (1993)	rats	lung elastin	14.4	0.8	6h/day; 79-90 days
43704_d	Rajini et al. (1993)	rats	lung injury	14.4	0.8	6h/day for 3 days
42369_e	Last (1991)†	rats	lung collagen	5	1 mg/m3 NaCl†	7 days continuous
42343_h	Schlesinger et al. (1990)	rabbits	eicosanoids (pulmonary)	3	0.3	2h exposure; 24 sacrifice hr
42343_i	Schlesinger et al. (1990)	rabbits	eicosanoids (pulmonary)	3	0.3	2h exposure; 0 sacrifice hr
42396_j	Schlesinger et al. (1991)	rabbits	eicosanoids (pulmonary)	3	0.3	2h exposure; 24 sacrifice hr
42239_k	Lee et al. (1989)	rats	enzymes; metabolizing	1.8	0.45	3 days continuous
15440_d	Bermudez (2001)	rats	DNA repair (lung cells)	1.2	0.3	3 days continuous

Table 2: First 7 variables of the FOREST data set

additivity	quotient	LCL	UCL	Reference	Species	Endpoint_Type	NO2	O3	Duration
Greater	0.4635	0.0514	1.5517	Reference	Species	Endpoint Type	NO2 (ppm)	O3 (ppm)	Duration
Greater	0.64	0.0298	4.1545	Reference	Species	Endpoint Type	NO2 (ppm)	O3 (ppm)	Duration
Greater	1.6380	.	.	Reference	Species	Endpoint Type	NO2 (ppm)	O3 (ppm)	Duration
Greater	4.8823	.	.	Reference	Species	Endpoint Type	NO2 (ppm)	O3 (ppm)	Duration
Greater	1.5348	.	.	Reference	Species	Endpoint Type	NO2 (ppm)	O3 (ppm)	Duration
Greater	9.9397	.	.	Reference	Species	Endpoint Type	NO2 (ppm)	O3 (ppm)	Duration
Less	-2.768	.	.	Reference	Species	Endpoint Type	NO2 (ppm)	O3 (ppm)	Duration
Less	-3.500	.	.	Reference	Species	Endpoint Type	NO2 (ppm)	O3 (ppm)	Duration
Less	-0.782	-3.044	-0.081	Reference	Species	Endpoint Type	NO2 (ppm)	O3 (ppm)	Duration
Less	-0.242	-0.344	-0.120	Reference	Species	Endpoint Type	NO2 (ppm)	O3 (ppm)	Duration
Greater	1.8655	.	.	Reference	Species	Endpoint Type	NO2 (ppm)	O3 (ppm)	Duration

Table 3: Next 10 variables of the FOREST data set

Finally, Output 1 shows a forest plot created with PROC SGPLOT. Title and footnote statements were used to create the titles and x-axis label. The Graph Template Language (GTL) was used to create a template with the marker colors and types, among other options, we wanted for the plot. Note that the arrows on some confidence intervals were added after plot creation with a graphics editor.



Output 1: Forest plot created with SGPLOT procedure.

CONCLUSION

In conclusion, the method presented here can be used to conduct an ANOVA in order to obtain comparisons not performed in an original analysis, when only summary statistics have been reported. The calculation of the quotient is useful for comparing data across experiments when different responses have been measured. We used this approach to compare health endpoints that can be grouped in categories (e.g., lung function), even if different metrics were used across studies. Finally the SGPLOT procedure is useful for combining point estimates and confidence intervals with descriptor information in order to create a detailed forest plot.

REFERENCES

Johns, DO; Stanek, LW; Walker, K; Benromdhane, S; Hubbell, B; Ross, M; Devlin, RB; Costa, DL; Greenbaum, DS. (2012). Practical advancement of multipollutant scientific and risk assessment approaches for ambient air pollution [Review]. Environ Health Perspect 120: 1238-1242. http://dx.doi.org/10.1289/ehp.1204939

Hidy, GM; Brook, JR; Demerjian, KL; Molina, LT; Pennell, WT; Scheffe, RD. (2011). Technical challenges of multipollutant air quality management. In GM Hidy; JR Brook; KL Demerjian; LT Molina; WT Pennell; RD Scheffe (Eds.). New York, NY: Springer. http://www.springer.com/environment/pollution+and+remediation/book/978-94-007-0303-2

Mauderly, JL; Burnett, RT; Castillejos, M; Ozkaynak, H; Samet, JM; Stieb, DM; Vedal, S; Wyzga, RE. (2010). Is the air pollution health research community prepared to support a multipollutant air quality management framework? Inhal Toxicol 22: 1-19. http://dx.doi.org/10.3109/08958371003793846

NRC (National Research Council). (2004). Air quality management in the United States. Washington, DC: National Academies Press. http://www.nap.edu/catalog.php?record_id=10728

Larson, DA. (1992). Analysis of variance with just summary statistics as input. Am Stat 46: 151-152. http://dx.doi.org/10.1080/00031305.1992.10475872

Lehman, RS. (1993). Analysis of variance with just summary statistics as input - Comment. Am Stat 47: 157-157.

Dilba, G; Bretz, F; Guiard, V. (2006). Simultaneous confidence sets and confidence intervals for multiple ratios. J Stat Plan Inference 136: 2640-2658. http://dx.doi.org/10.1016/j.jspi.2004.11.009

ACKNOWLEDGEMENTS

The authors would like to acknowledge the other members of the Toxicological Effects of Criteria Pollutant Mixtures project team:

Brianna Young, University of North Carolina
Adrien Wilkie, University of North Carolina
Meagan Madden, Duke University
Lindsay Stanek, U.S. EPA
Doug Johns, U.S. CDC

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Laura Datko Williams, ORISE Fellow at U.S. EPA
Mailing Address:
U.S. EPA
Attn: Laura Datko Williams
Mail Drop B243-01
Research Triangle Park, NC 27711

(919) 541 - 0025
datko-williams.laura@epa.gov