Los Alamos
NATIONAL LABORATORY

memorandum
Environment, Safety, and Health Division
ESH-17 Air Quality Group

To/MS: Distribution

Thru: Craig Eberhart, ESH-17, J978

From/MS: D.J. Dawson, ESH-17, J978

Phone/FAX: 5-3850/5-8858

Symbol: ESH-17:99-146

Date: July 19, 2000

TRITIUM OUTLIER CALCULATIONS

I have completed the tritium outlier calculations, which are attached to this memo. When I started these calculations, I used a program called DataQuest, which was designed by the EPA to calculate basic statistics in environmental studies. Apparently this program does not always accurately determine the correct number of outliers even though it does accurately determine the mean, standard deviation, and other statistical information.

The first table labeled "Software Outlier Calculations" shows the results of running the Rosner's test with the p-value set at .01 or .05. It then shows the mean ("x") and mean plus three standard deviations ("x+3s"). The column labeled "# high" represents the number of the outliers given by the program that are from the high values if the number is not equal to the total number of outliers. The table labeled "Excel Outlier Calculations" shows comments on both the results of the spreadsheet calculations and the number of outliers according to the DataQuest program. As seen, sometimes there is a very substantial difference. Most of the time the spreadsheet calculations contained more outliers, however, there were five times that DataQuest found more.

In the "Excel Outlier Calculations" table, the "x" represents the mean with all the outliers taken out, the "x+3s" represents the mean plus 3 standard deviations with all the outliers taken out, and the "Min. Outlier" represents the last outlier taken out. The "Quest Outlier" represents the number of outliers that DataQuest found that, while the "# high" category shows how many of the DataQuest outliers that from the high end of the points. In the comment field, there are a wide variety of comments, which include the following:

> "Both same"- Both the manual calculations and DataQuest came up with identical answers.
>
> "Computer fail to find some"- DataQuest did not find all the outliers that my calculations did.
>
> "Computer greatly underestimated"- DataQuest only found 2 outliers, when my calculations showed at least 10 outliers.
>
> "Quest not able to do more"- DataQuest reached its limit of 5 when there were more outliers by my calculations.
>
> "???Takes low value"- Unsure how DataQuest was able to take out the low value. The low value was too close to the mean to be selected as the extreme value.
>
> "Unsure how quest got x"- DataQuest found more outliers than I did. Most of the time it took out some low-end values, which were questionable and determined more outliers than my calculations.

To evaluate these potential action levels, I propose a meeting (April 22 or 23) to discuss their usefulness. If there are any comments on the data before the meeting occurs (such as changes needed, more information needed, etc.) feel free to talk to me about them. Also, the radionuclide data should be done in the next day or two as well using the +3s method and Rosner's method. I will have manually calculated all the action levels using the Rosner's method as a result of the DataQuest's problems. So when it is done, I shall send it around, as well, for review before the meeting.

DJD/db

Att: a/s

Distribution:
Doug Stavert, ESH-17, J978
Jean Dewart, ESH-17, J978
Dave Kraig, ESH-17, J978
Scott Miller, ESH-17, J978
Ernie Gladney, ESH-17, J978

Cy:
ESH-17 File