UNITED STATES ENVIRONMENTAL PROTECTION AGENCY

National Center for Environmental Assessment
Research Triangle Park, NC 27711

June 5, 2002

Recipients of Third Draft
EPA PM Criteria Document

Dear Colleagues:

The U.S. EPA recently released the two-volume Third External Review Draft of the EPA document "Air Quality Criteria for Particulate Matter", for a 60-day public comment period (ending July 10, 2002) and for review by the Clean Air Scientific Advisory Committee (CASAC) at a public meeting scheduled for July 18-19, 2002 at EPA facilities in Research Triangle Park, NC. That draft EPA PM Air Quality Criteria Document (PM AQCD; EPA/600P-99/002aC and EPA/600/P-99/002bC) assesses newly available information on health and ecological effects of exposures to ambient air PM, to provide key scientific bases to support the current periodic review of U.S. Particulate Matter National Ambient Air Quality Standards (PM NAAQS). Chapter 8 (Epidemiology) of that draft document assesses numerous time-series studies and other types of PM epidemiology studies; and Chapter 9 (Integrative Synthesis) integrates information from the epidemiologic chapter with information from other chapters.

The U.S. EPA was recently informed by the Health Effects Institute (HEI) of a generally unappreciated aspect in the use of S-Plus statistical software often employed to fit generalized additive models (GAM) to data in time-series analyses. Salient points regarding this issue are summarized in this letter. More detailed information about the overall S-Plus/GAM issue and preliminary findings from initial reanalyses of National Morbidity Mortality and Air Pollution Studies (NMMAPS) can be found in the HEI letter informing U.S. EPA and other HEI sponsors about the issue, as appended here and/or posted on the HEI website (http://www.healtheffects.org). In its letter, HEI also outlines steps they plan to take to ensure timely peer-review and open discussion of new findings emerging from the reanalyses of the NMMAPS data.

As indicated in the HEI letter, in estimating the GAM, the S-Plus program uses an iterative process that comes to completion when the improvement in model fit is less than a preset criteria. Investigators at Johns Hopkins University (J. Samet and colleagues) used the S-Plus default convergence criteria in conducting HEI-funded PM epidemiology analyses reported in their published NMMAPS multi-city studies discussed in Chapter 8 of the Third Draft PM AQCD. The NMMAPS investigators found that, for a given city, the default convergence criteria in S-Plus appear to be inadequate to assure that convergence of its iterative estimation procedure actually reaches the maximum likelihood value. Depending on the city, use of the default

convergence criteria can bias the estimate of relative risk of air pollution upwards or downwards. Initial analyses suggest that these changes may be most acute when the effect to be estimated is relatively small or when substantial colinearity is present.
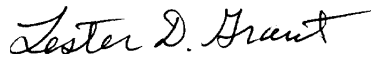
The HEI letter also notes that the NMMAPS investigators have begun to address this issue using Generalized Linear Models (GLM) and have tested these techniques to ensure that they do not share the same analytical issue. Preliminary reanalyses of the NMMAPS 90-city dataset, using GLM methods, suggest that the individual city effects are sometimes larger, but more often smaller than those previously reported. On average, the size of the new city-specific estimate varies by 23% from the original value, and the original pooled estimates of effects for 90 cities appear to be reduced by use of the newly employed analytic techniques. As an example, for total mortality at lag 1, the updated value is approximately 0.2% per 10 $\mu g/m^3$ PM_{10} (about half of the original value of 0.4% per 10 $\mu g/m^3$ PM_{10} for the NMMAPS nationwide effect estimate discussed in the PM AQCD). The new PM effect estimate, while smaller, is still statistically significant and thus far apparently not sensitive to adjustments for other gaseous pollutants (O_3, CO, NO_2, and SO_2). In summary, these new analyses suggest that the new findings, though resulting in smaller estimates of effect, appear to be similar to the original in many respects, and the full extent of any differences resulting from these new analyses are still being explored.

Obviously, the above-noted developments may have implications for other time-series studies (both of air pollution and other topics) which used the GAM technique in the S-Plus software. The effects of using appropriate modifications of the default convergence criteria code in the S-Plus software or other alternative techniques cannot now be predicted; new estimates of effects for any specific location(s) may possibly increase or decrease from values reported in the published literature for those PM studies that employed the default convergence criteria in S-Plus analyses involving GAM techniques. Efforts are underway by U.S. EPA staff (1) to identify which of the published studies assessed in Chapter 8 of the Third Draft PM AQCD may have employed such techniques in S-Plus analyses, (2) to inform authors of such studies about the developments discussed above, and (3) to develop plans and projected schedules for appropriate revision of Chapters 8 & 9 of the draft PM AQCD to adequately address this issue.

In the meantime, as per discussion with CASAC Chair, Dr. Phil Hopke, the U.S. EPA plans to proceed with the current review of the Third Draft PM AQCD. This includes acceptance of written public comments on that draft document to be submitted by July 10, 2002. Holding the July 18-19 CASAC meeting is also viewed by EPA as still being very useful, both (1) to hear comments and allow for review of most chapters in the Third Draft PM AQCD, and (2) to discuss with CASAC the above-noted developments, the status of EPA efforts to identify potentially affected PM epidemiology studies, and plans for addressing this and related issues through further revisions of draft PM AQCD Chapters 8 & 9. Please also note that it is U.S. EPA's view that, while closure on Chapters 8 & 9 of the Third Draft PM AQCD will not be possible at the July CASAC meeting, the Agency would nevertheless benefit from receipt of public comments and preliminary CASAC review of Chapters 8 and 9, to assist in shaping later revisions to those chapters.

We look forward to working with CASAC, HEI, the NMMAPS investigators, and other members of the scientific community to better understand the scope of issues raised above and how best to address them in order to allow expeditious completion of the PM AQCD review. The need to address this issue will also impact completion of the next draft PM Staff Paper by EPA's Office of Air Quality Planning and Standards (OAQPS). This draft Staff Paper was to be reviewed by CASAC at a September 18-19, 2002 meeting, that will now be rescheduled following the July 2002 CASAC meeting on the draft PM AQCD.

Sincerely,

Lester D. Grant, Ph.D.
Director, NCEA/RTP

May 30, 2002

Dear Colleague:

I write to provide you with advance notice of some valuable new insights developed and communicated to HEI in the last few weeks by the investigators of the National Morbidity, Mortality, and Air Pollution Study (NMMAPS) at Johns Hopkins University. As you recall, NMMAPS included the first nationwide, systematic analyses of air pollution, weather, and mortality, in the 90 largest cities in the United States. It has provided, and continues to provide, a unique database and new methods for investigating these relationships, and their possible variation by geographic region, season and other factors.

In the interest of continuing to advance the science and the statistical techniques of NMMAPS, the investigators at Johns Hopkins have been exploring their original results, testing them against different assumptions, and examining the methods they used. In that process, they have identified a generally unknown aspect of the S-plus statistical software that they and many others have used to fit generalized additive models (GAM) to the data in these time-series analyses. In estimating the GAM, the S-plus program has an iterative process that comes to completion when preset criteria (referred to as the convergence criteria) are met. The default criteria in S-plus were used by the NMMAPS investigators. The NMMAPS investigators have recently found that for a given city the default "convergence criteria" in S-plus appear to be inadequate to assure that the convergence of its iterative estimation procedure actually reaches the maximum likelihood value. Depending on the city, and the trends of air pollution, mortality and weather in that city, use of the default criteria can bias the estimate of relative risk of air pollution upwards or downwards. Initial analyses suggest that these changes may be most acute when the effect to be estimated is relatively small. Overall, these changes would also affect the pooled relative risks developed by NMMAPS and any other similar studies.

To their credit, the NMMAPS scientists have taken the initiative to investigate alternative analysis techniques to address this issue using Generalized Linear Models (GLMs) and have tested these techniques to ensure that they do not share the same analytical issue. Using GLM, they have begun to systematically revisit all of their analyses to examine the implications for their results. Redoing all of the major prior analyses requires the team to fit thousands of regression models; in analyses already completed more than 3,000 models have been refit and more are underway. HEI, as would be its normal procedure, has convened the NMMAPS Review Panel of the HEI Review Committee, chaired by Dr. Sverre Vedal of the National Jewish Medical and Research Center in Denver, to begin reviewing these results. Given the importance of the NMMAPS study in the current ongoing use of its results in the CASAC and other standard-setting processes, HEI is working to report on that review in the most timely manner possible.

Although these new analyses by the NMMAPS investigators have not been completed nor reviewed yet by HEI, we want to be sure that any results from such analyses are made available to all interested parties at the earliest possible time. To that end, the NMMAPS investigators have reported to HEI several preliminary findings that we want to share with you at this time.

First, the individual city effects in the new analyses are sometimes larger, and more often smaller, than previously reported. On average the size of the new city-specific estimate varies by 23% from the original value. See Figure 1, a scatter plot of the new estimates against the old estimates for the 90 cities.

Second, the pooled estimates of effects using their new analysis techniques for 90 cities appear to be lower than found in the original NMMAPS reports. For example, for total mortality at lag 1, the updated value is around 0.2% per 10 micrograms per cubic meter, as compared to being around 0.4% per 10 micrograms per cubic meter in the original analyses. Figure 2 displays the new and old pooled relative risk estimates for total mortality and PM_{10} at lags 0, 1 and 2 days.

Third, the investigators have re-done the 90-city analysis of cause-specific mortality. Although the pooled relative risks are all smaller, it appears that the greatest relative risk estimate continues to be for cardiorespiratory mortality. Figure 3 presents the posterior distributions for the pooled effect of lag 1 PM_{10} on cause-specific mortality.

Fourth, these initial analyses suggest that these smaller PM_{10} relative risk estimates appear to not be sensitive to adjustment for other gaseous pollutants including O_3, CO, NO_2 and SO_2. Figure 4 shows the posterior distributions for the PM_{10} relative risk at lag 1 on total mortality for an unadjusted model and for several multi-pollutant models. Further analyses, including those that test the relative risks of the other pollutants, and then see how they are affected by the inclusion of PM in the model, are underway.

In summary, the NMMAPS team has begun to report promptly on an improved methodology that results in a smaller pooled estimate of the relative risk. These preliminary analyses suggest that the new findings, though resulting in smaller estimates of effect, appear to be similar to the original in several respects. Obviously, the full extent of any differences resulting from these new analyses is still being explored, and the NMMAPS Review Panel (and the broader HEI Review Committee of which its work is a part) has yet to review and comment on the significance of these findings. In the interest of ensuring the open and transparent discussion of these valuable new findings, HEI plans to pursue the following:

- Systematic revisiting by the investigators, and HEI Review Committee examination, of all of the relevant NMMAPS analyses.

- Periodic updates on progress, and preparation of a more comprehensive presentation of reviewed results at the July 18-19 meeting of CASAC, and

- A report of findings and a commentary from the HEI Review Committee by Fall 2002.

Beyond NMMAPS, this discovery by the NMMAPS team may have implications for a number of other time-series studies, both of air pollution and other topics, which used the GAM technique in the S-plus statistical software. It is not possible to predict the effects of any changes

in other studies; as with the NMMAPS cities, the estimates of effects in any particular location(s) may change upwards or downwards. HEI has begun the process of identifying any of its own studies that may be affected, and the NMMAPS investigators have begun informing the broader scientific community of their preliminary findings so that other investigators can explore what, if any, implications these findings might have.

Should you have any questions concerning these findings, please feel free to contact us.
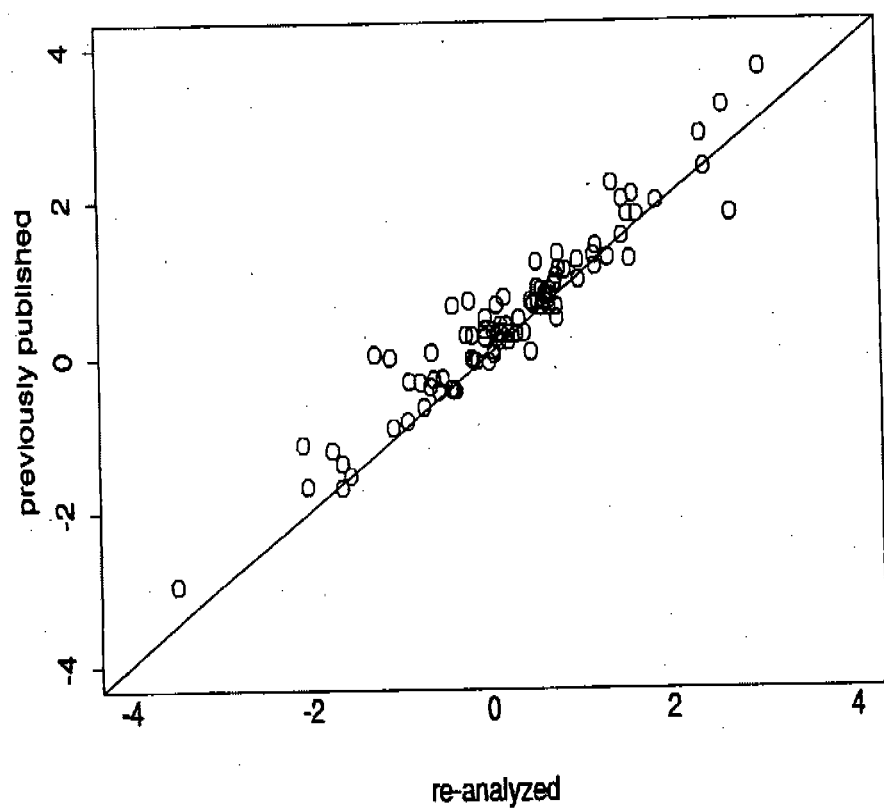
Sincerely,

Dan Greenbaum
President

Figure 1: Relative rates of total mortality at lag 1: re-analyzed versus previously published estimates, 90 U.S. cities.

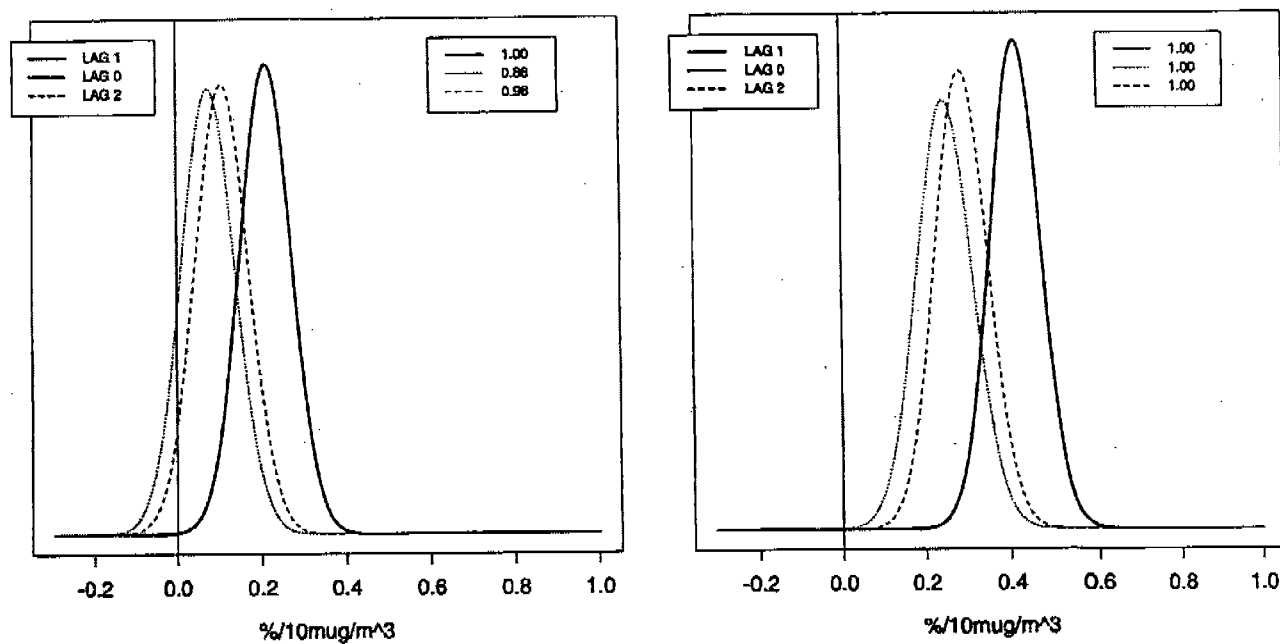Re-analyzed pooled estimates   Previously published pooled estimates

Figure 2: Left Panel: Marginal posterior distributions for the **re-analyzed** pooled effects of PM_{10} on total mortality at lag 0,1, and 2, for the 90 U.S. cities. The box at the top right provides the posterior probabilities that the overall effects are greater than 0. Right Panel: Marginal posterior distributions for the **previously published** pooled effects of PM_{10} on total mortality at lag 0,1, and 2, for the 90 U.S. cities. The box at the top right provides the posterior probabilities that the overall effects are greater than 0.
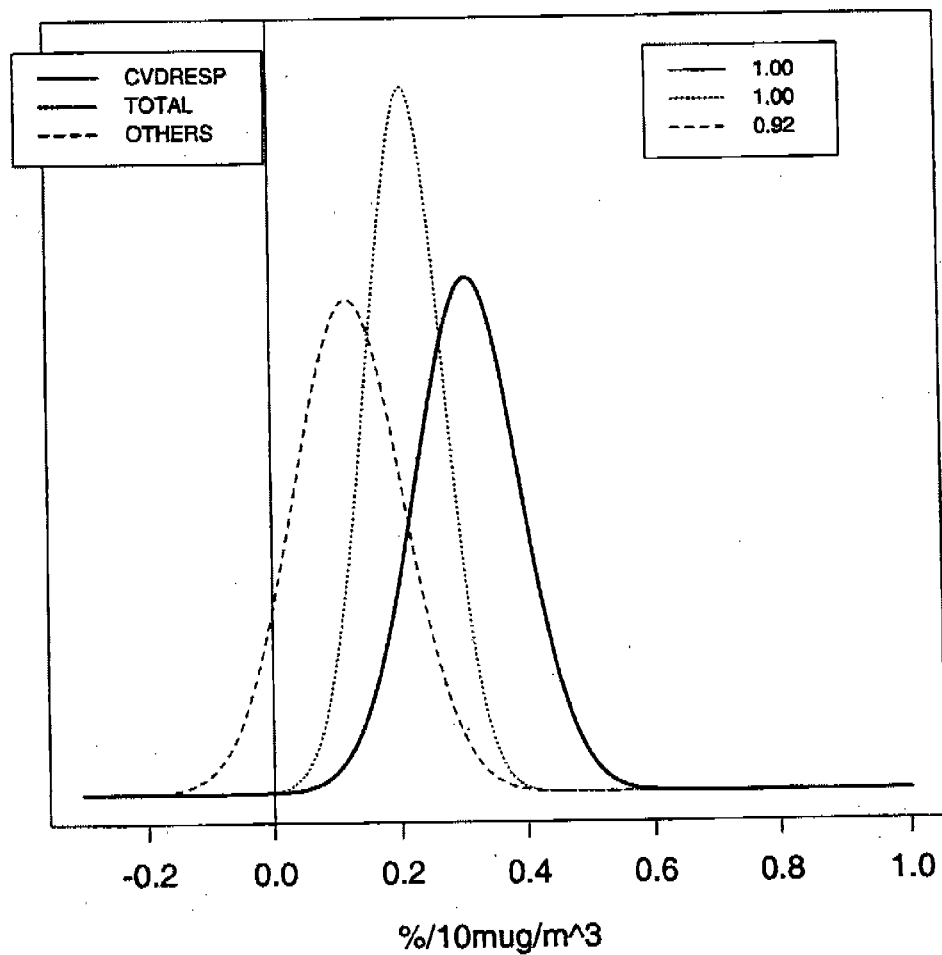
Figure 3: Marginal posterior distributions for the **re-analyzed** pooled effects of PM_{10} at lag 1 for total mortality, cardiovascular-respiratory mortality and other causes mortality, for the 90 U.S. cities. The box at the top right provides the posterior probabilities that the overall effects are greater than 0.
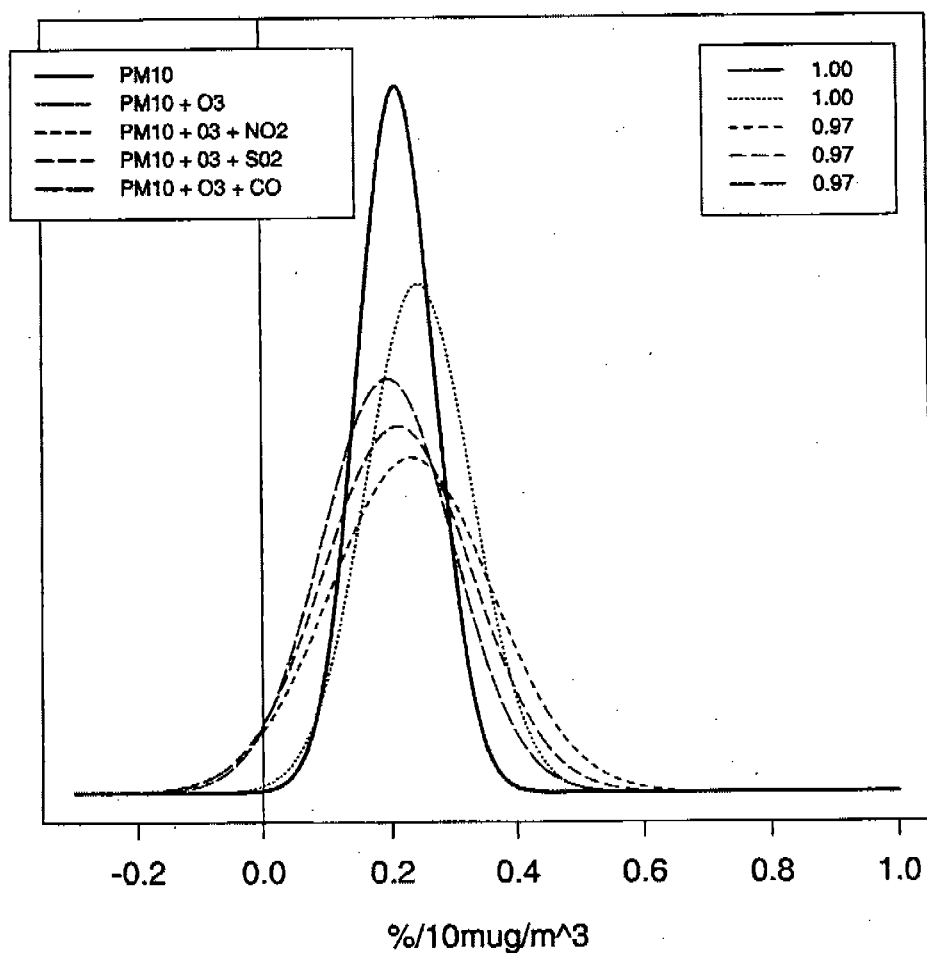
Figure 4: Marginal posterior distributions for the re-analyzed pooled effects of PM_{10} on total mortality at lag 1 with and without control for other pollutants, for the 90 U.S. cities. The box at the top right provides the posterior probabilities that the overall effects are greater than 0.