Overview and Summary of Basis for Short-Term Mortality Estimates in EPA's Staff Paper

EPA's *Staff Paper* contains estimates of changes in two types of mortality risks that might result from the alternative standards for $PM_{2.5}$: "short-term" and "long-term" mortality. This paper focuses on the short-term category of mortality risk related specifically to $PM_{2.5}$ concentrations.

Overview of Short-Term Mortality Risk Estimation Methods and Issues

Short-term mortality is a term coined to refer to possible risks of dying prematurely as a result of a brief (*i.e.*, "short-term") period of exposure to elevated concentrations of an air pollutant such as $PM_{2.5}$. It is consistent with the concept of acute dose and response, as contrasted responses to chronic doses over an extended period of time.

Epidemiological studies that can assess short-term $PM_{2.5}$ mortality risks are done by studying how much the death rate fluctuates from day to day, and whether days observed to have a higher-than-average number of deaths tend to coincide with days that have elevated $PM_{2.5}$. Because daily $PM_{2.5}$ fluctuations occur on a relatively local scale, short-term studies tend to be conducted for individual cities. (Even when multiple cities are included in a study, the fundamental assessment of correlation between $PM_{2.5}$ and mortality risk is performed city by city within the study. City-specific results might be aggregated into a single estimate as a second step of the study, but there is always an underlying city-specific relationship.)

Statistical methods to assess the $PM_{2.5}$-associated portion of risk attempt to "control" for the myriad other important determinants of death. If all the key causal factors are not properly controlled for, estimates of the risk imposed by $PM_{2.5}$ may be biased, and thus unreliable for either determining if an effect exists, or for quantifying the magnitude of that effect. Following are some of the common forms of controls that are included:

- Usually only non-accidental forms of death are counted, as no one has hypothesized that $PM_{2.5}$ might contribute to any of the accidental forms of death. If $PM_{2.5}$ exposures were known to affect health risk in a particular manner, then it would be appropriate to study only the sub-category of non-accidental deaths associated with that mode of physical impact, in order to further eliminate extraneous causes of variability in daily death rates. Researchers often estimate $PM_{2.5}$ risks for various sub-categories of non-accidental death, such as cardiovascular deaths, or deaths from respiratory illness. However, until a $PM_{2.5}$-mediated mechanism of impact is established, such efforts are as much an attempt to search for evidence of a pathway of risk as they are an effort to provide better statistical controls.

- Death rates vary on a temporal basis. Stable patterns are known to exist on a seasonal and even day-of-week basis. These patterns may have some linkage to weather, but they also exist regardless of a day's specific weather outcome. Researchers attempt to control for the non-weather component of this pattern by various methods of adding terms to the statistical relationship that provide "temporal smoothing." This is an area of great uncertainty in short-term $PM_{2.5}$ studies because there is no objectively "correct" method

for temporal smoothing, and the most appropriate method, if it could be identified, might be different for each city. This is problematic because researchers have increasingly realized (particularly since the period of "reanalyses" in 2002-2003) that the size of the estimated $PM_{2.5}$ effect, and whether or not it is statistically significant, can often vary greatly depending on how much temporal smoothing has been accomplished in the analysis.

- Weather, particularly weather extremes, is known to affect mortality risk even after accounting for temporal trends. The proper way to control for weather effects is also subject to uncertainty and analyst judgment. The most common forms of weather variables used include temperature (usually as deviations from a neutral temperature such as 72° F) and relative humidity.

- Other pollutants co-vary along with $PM_{2.5}$ from day to day, and none of them can be clearly exonerated of any causal role in observed mortality risk. If controls are not properly applied for other pollutants that have a causal impact, the statistical analysis could assign an association to $PM_{2.5}$ that may be partly or totally due to a different culprit pollutant. It is, however, extremely difficult to apply such controls, because pollutants are often highly correlated with each other in time, and this can make it very difficult or impossible for statistical methods to identify the separate roles of each. However, in general one can expect that a $PM_{2.5}$ effect that has been estimated with $PM_{2.5}$ as the only pollutant in the statistical model (*i.e.*, a 1-pollutant formulation) is more likely to be biased than a $PM_{2.5}$ effect that has been estimated while controlling for other potential culprit pollutants (*i.e.*, in a 2-pollutant or multi-pollutant formulation).

- Effects lags may exist, in that the response to a pollutant (or other variable) may become evident in the mortality data several days after the elevated exposure level. Without any knowledge of a mechanism of effect, there is no *a priori* reason to expect any particular length of lag. The most common method has been to consider a range of lags, from 0 to as many as 6 days. However, if another pollutant is the culprit, and it is not controlled for, *or its lag is not correctly specified* in a multi-pollutant formulation, $PM_{2.5}$ might pick up that other pollutant's effect with some lag periods but not others. Relying on a result for the lag that produces the most prominent $PM_{2.5}$ effect in each study, and allowing the length of lag used to thereby vary with the study being used, might result in an unwarranted overstatement of the effect attributable to $PM_{2.5}$.

Another complication in the case of $PM_{2.5}$ risk estimation relates to the complexity of $PM_{2.5}$ itself. This "pollutant" is made up of very many and diverse chemical and physical forms. Their only shared attribute is that they are all present in the air in solid or liquid droplets form ("particles") rather than as gases, and that those particles all are smaller than 2.5 microns in aerodynamic diameter. Therefore, even if one were to have confidence that a $PM_{2.5}$ health effect has been detected by an epidemiological study (and that it is not a proxy for some other pollutant or other non-pollutant factor), one has no idea which specific constituents in the mix are the culprits. Only a few epidemiological studies have attempted to explore the role of individual constituents comprehensively, and none of these is being used in EPA's risk analyses. This means that any estimate of mortality risk reduction from alternative standards has inherently large uncertainty, that will always encompass "no change" at all, because there is no way of

knowing if the culprit constituent will be reduced as a result of local efforts to attain an air quality standard for $PM_{2.5}$.

Sound controlling is essential to sound statistical estimation, and it is clear that controlling is a difficult and complex enterprise for $PM_{2.5}$ that has led to the use of quite sophisticated statistical tools. One of the advanced statistical methods that has been used, Generalized Additive Models (GAM), led to substantial controversy because it was determined that many researchers had not using the tool properly, and a large fraction of the studies available as of 2001 had to be reanalyzed. In the ensuing process of reanalysis, other issues came to the forefront. First, the sensitivity of $PM_{2.5}$ risk estimates to the degree of spatial smoothing emerged as a new concern that has yet to be resolved. Second, it became known that GAM-based estimates of standard errors were biased low, which would have the effect of overstating the degree of significance of $PM_{2.5}$ risk estimates. Even though GAM-based estimates that EPA is now using have all been corrected through reanalyses, it should be kept in mind that EPA's continued reliance on GAM methods may lead to too many results being categorized as statistically significant. As will be seen, even with this potential bias in the evidentiary basis, a very large fraction of the available $PM_{2.5}$ short-term risk estimates are statistically insignificant.

Short-Term Mortality Risk Estimates in EPA's *Staff Paper*

EPA's *Staff Paper* presents estimates of short-term mortality changes for each of eight individual cities: Boston, Detroit, Los Angeles, Philadelphia, Phoenix, Pittsburgh, St. Louis, and San Jose. These cities are not considered representative. Rather, they reflect the paucity of available evidence on short-term $PM_{2.5}$ mortality risks: these are the only cities that met several criteria for data availability, including the existence of a $PM_{2.5}$ daily mortality study and sufficient $PM_{2.5}$ air concentration data.

For each of the eight cities, the risk estimates at alternative standards reported in the *Staff Paper* are based on a single regression from a single paper. Each of those papers, however, contains multiple other risk estimates, not all of which are consistent with the estimate EPA has elected to use. For some of the cities, there are other epidemiological papers available that provide different conclusions about $PM_{2.5}$ mortality risk, usually based on the same underlying air quality and mortality data. Thus, EPA's *Staff Paper* does not provide a comprehensive summary of the variance in the epidemiological evidence.

Initially, EPA provides the statistical confidence intervals of those single estimates (*e.g.*, in Chapter 4 of the *Staff Paper*). However, even this minimal reflection of the variance in the underlying epidemiological data is dropped by the time the results are presenting in Chapter 5 of the *Staff Paper*.

Attached to this overview is a set of 1-page summaries, one for each city in the short-term mortality risk assessment. Each one identifies what EPA is assuming for the single short-term mortality risk estimate for that city, along with a more complete description of the full body of published epidemiological evidence. Several themes appear across the set of these 1-page summaries:

- EPA is consistently using 1-pollutant formulations, although multi-pollutant formulations frequently diminish the strength of evidence for a $PM_{2.5}$ mortality association.
- EPA is consistently using estimates that find the largest effect, even when sensitivity analyses on temporal smoothing exist that diminish the strength of evidence for that $PM_{2.5}$ mortality association.
- When a study finds a significant effect for one lag period and an insignificant effect for a lag period that is one day longer or shorter, EPA uses the significant one. Thus, the lag period used in each city varies in the Risk Assessment, while obscuring the overall non-robustness of the effects estimates in all the cities.
- EPA is using $PM_{2.5}$ risk estimates from studies that in fact concluded that if there is any pollutant-health association at all, it is one of the gaseous pollutants.
- EPA consistently relies on GAM-based estimates even when alternative estimates are available that use linear methods for temporal smoothing (*e.g.*, "splines" or "GLM") that have more reliable ability to assess an estimate's significance level.

Table 1 summarizes the overall findings for each of the eight cities. The details supporting the information in Table 1 can be found in each of the one-page summaries.

Table 1. Summary of Evidence on $PM_{2.5}$ Association with Short-Term Mortality for Eight Cities in EPA's Risk Analysis

City	Is the estimate EPA uses statistically significant?	Percent of all available estimates for this city that are significant	General synopsis of findings for this city
Boston	Yes	80%	*Evidence of a $PM_{2.5}$ mortality risk in Boston is subject to uncertainty related to how to properly provide statistical controls, and it has not been explored in a multi-pollutant manner at all. The magnitude of the effect, even if causal, is overstated by the single estimate EPA has used.*
Detroit	No	0%	*There is no evidence of a $PM_{2.5}$ mortality risk in Detroit that is statistically significant, even without consideration of the role of other pollutants in multi-pollutant formulations. The authors have identified important concerns about how to apply statistical controls that undermine the meaningfulness of even the insignificant $PM_{2.5}$ risk estimate that EPA has chosen to use.*
Los Angeles	No	9%	*This paper finds evidence <u>against</u> assuming a causal role of $PM_{2.5}$ in any of the correlations between $PM_{2.5}$ and mortality in Los Angeles. Even though the risk estimate that EPA is using from this paper is relatively small and is insignificant, it is still being used out of context, and in a manner directly inconsistent with the author's own interpretation.*
Philadelphia	Yes	47%	*This paper finds evidence <u>against</u> assuming a causal role of $PM_{2.5}$ in any of the correlations between $PM_{2.5}$ and mortality in Philadelphia. The relatively large and significant risk estimate that EPA is using from this paper has been taken out of context, and is being used in a manner directly inconsistent with the authors' own interpretation.*
Phoenix	Yes	30%	*The study EPA is using provides only very weak support for inferring a $PM_{2.5}$ effect in Phoenix, and it is further weakened by the findings in two other papers that EPA has not used*
Pittsburgh	No	0%	*This paper does not find any evidence of a significant association between $PM_{2.5}$ and mortality in Pittsburgh. Use of any risk estimates from this paper to indicate mortality benefits from reduced $PM_{2.5}$ is inconsistent with its overall findings.*
St. Louis	Yes	36%	*Evidence of a $PM_{2.5}$ mortality risk in St. Louis is subject to uncertainty related to how to properly provide statistical controls, and it has not been explored in a multi-pollutant manner at all. The magnitude of the effect, even if causal, is highly overstated by the single estimate EPA has used.*
San Jose	Yes	57%	*This paper finds more evidence of an effect than any of the others used in EPA's Risk Analysis, but even in this study the evidence is not very robust or coherent. Additionally, the rapidly changing air quality in San Jose over the study period raises questions (e.g., about thresholds) that the paper does not acknowledge.*