Prepared Statement of
Anne E. Smith, Ph.D.
at a Hearing on
The American Energy Initiative
– A Focus on EPA's New Proposal to Tighten National Standards for Fine Particulate Matter in the Ambient Air –
by the
Subcommittee on Energy and Power
Energy and Commerce Committee
United States House of Representatives
Washington, DC

June 28, 2012

Mr. Chairman and Members of the Committee:

Thank you for your invitation to participate in today's hearing. I am Anne E. Smith, a Senior Vice President of NERA Economic Consulting. I am a specialist in environmental risk assessment and integrated assessment to support environmental policy decisions, which was a core element of my Ph.D. thesis at Stanford University in economics, with a minor concentration in decision sciences. I have performed work in the area of air quality cost and benefits analysis and risk assessment over the past thirty years, including as an economist in the USEPA's Office of Policy, Planning, and Evaluation, as a consultant to the USEPA Air Office, and in many consulting engagements since then for government and private sector clients globally. I have also served as a member of several committees of the National Academy of Sciences focusing on risk assessment and risk-based decision making. I have analyzed costs, risks and benefits of many U.S. air policies, including fine particulate matter ($PM_{2.5}$), regional haze, ozone, mercury, NO_2, SO_2, and greenhouse gases. I have been extensively involved in assessment of the evidence on risks from ambient $PM_{2.5}$ since EPA first turned to the task of identifying an appropriate National Ambient Air Quality Standard (NAAQS) for $PM_{2.5}$ over fifteen years ago. I have also been active in the assessment of visibility impacts since twenty years ago, when I supported the Grand Canyon Visibility Transport Commission's integrated analysis of western regional haze management options.

I thank you for the opportunity to share my perspective today on the EPA's new proposal to tighten the national ambient air quality standards (NAAQS) for fine particulate matter ($PM_{2.5}$). My written and oral testimonies reflect my own opinions, and do not represent any position of my company, NERA Economic Consulting or of any of its clients.

Executive Summary

The focus of this hearing is EPA's new proposal to tighten the $PM_{2.5}$ NAAQS. Several critical features of the new proposal are worth highlighting because they merit much closer inspection and understanding. As a group, these features suggest that the proposed PM NAAQS could be more costly than EPA's cost estimate, and its benefits are far more uncertain than EPA's benefits analysis indicates.

Two features of the Proposed Rule that could result in costs higher than EPA has estimated are:

1. EPA proposes to require a new set of monitors placed near roads in each air quality area, while simultaneously proposing to eliminate determination of attainment with a spatial average of $PM_{2.5}$ from all monitors in an area, so that attainment will be determined by each area's single worst-case monitor. Given

that it is highly likely that most near-road monitors will have higher $PM_{2.5}$ readings than community-oriented monitors, this change is likely to make the proposed NAAQS much more stringent than EPA has estimated. EPA has not attempted to characterize how much *more* stringent the standard will become as a result of these two changes to the NAAQS, or its cost implications. (See **Section 1** for further explanation.)

2. EPA proposes for the first time a secondary standard specifically to limit urban visibility degradation. The proposed visibility secondary standard would not be set equal to the primary standards, but would be set using a highly complex and arcane indicator called the "deciview." The following two facts about the deciview are generally unknown to all but a few "visibility experts":
   (a) seemingly small changes in the deciview level imply much larger changes in ambient $PM_{2.5}$ concentrations; and (b) a uniform national deciview standard will limit ambient $PM_{2.5}$ concentrations to very different levels in different cities. That is, a deciview-based standard implies a highly non-uniform $PM_{2.5}$ NAAQS.
   These two facts imply substantial uncertainty about the cost of the Proposed Rule if even seemingly slight changes are made in the proposed form and level of the proposed visibility standard. (See **Section 2** for further explanation.)

Perhaps most important of all with respect to the proposed visibility standard is that EPA has developed its case for the need for that standard using a type of study that has been demonstrated to be unreliable for determining the amount of visibility degradation that can be said to adversely affect the public welfare. Thus, in return for the planning complexity and cost uncertainty that the visibility standard would create, there would be no scientific basis for believing it will reduce adverse effects on the public welfare. Indeed, EPA's method cannot even credibly establish that urban visibility is causing adverse public welfare impacts under the current PM NAAQS. (See **Section 3** for further explanation.)

EPA's estimates of the health benefits of the proposed PM NAAQS are also far more uncertain than EPA admits. The Administrator's own rationale for setting the primary annual standard at a level no lower than 12 $\mu g/m^3$ is based on the greatly increasing uncertainty that the Administrator expresses regarding EPA's own projections of health benefits from yet-lower $PM_{2.5}$ levels. However, none of these admitted uncertainties are reflected in EPA's estimates of the benefits of the Proposed Rule. Uncertainty prevents the Administrator from setting the standard lower, but this hasn't prevented EPA from taking full credit for such uncertain benefits. The benefits analysis is thus inconsistent with the rationale for the standard. (See **Section 4** for further explanation.)

Further, almost all of EPA's estimate of benefits is predicated on a presumption that the statistical ("epidemiological") associations between chronic ambient $PM_{2.5}$ concentrations and mortality risk are causal in nature, and that all $PM_{2.5}$ constituents are equally potent. Even the presumption of causality is still subject to question, as has been demonstrated by a $PM_{2.5}$ chronic risk study published in 2011. (See **Section 5** for more explanation.) Uncertainty about the causality presumption means there is a possibility that there will be no benefits at all from a tightened $PM_{2.5}$ NAAQS. Causality uncertainty, which affects benefits estimates from $PM_{2.5}$ reductions at all levels (even benefits from reductions above the current standard), also is not reflected in EPA's benefits estimates. EPA's science assessment, which is the source of EPA's assertion that the chronic mortality risk associations are causal, was written before the 2011 paper was published. When strong contradictory evidence emerges, as is the case now, it would seem prudent public policy to more closely examine and deliberate the new evidence before locking into major and costly new regulatory actions.

The Proposed Rule was released on June 14, 2012. As of the time of this writing (June 27), EPA has not released its Regulatory Impact Analysis (RIA). Only a very brief summary of the RIA results is provided in the Proposed Rule and the associated Fact Sheet. Lack of access to much of the underlying analysis for EPA's cost and benefit estimates is frustrating because the above issues should be studied in greater depth, and that cannot be done without more backup to EPA's calculations of benefits and costs. Even without the relevant documentation, however, the basic points I have identified above provide concerns that the PM NAAQS rule will likely cost more than EPA has estimated, while providing benefits that are much more uncertain than EPA's predictions. Benefits uncertainty includes the possibility of no chronic mortality benefits, which account for over 90% of the value in EPA's benefits estimate.

Section 1. Changes in Monitoring and Attainment Requirements Make the Proposed Standard Tighter than EPA Estimates, and Likely More Costly to Meet

Some of the Proposed Rule's changes to the PM NAAQS relate to monitoring and attainment determinations. For one, the Proposed Rule would require a new set of $PM_{2.5}$ monitors to be placed near roads in each air quality area. Addition of such monitors to the monitoring network will help provide information about how much higher $PM_{2.5}$ concentrations may be in the vicinity of traffic. At present there is only limited evidence, but it suggests that the near-road concentrations could be substantially higher than at the standard community-based monitors presently being used to assess attainment.[1] Thus, near-road monitors can reasonably be expected to become the worst-case monitors for each air quality area.

Simultaneously, the Proposed Rule also would eliminate the ability to determine attainment using the spatially-averaged concentration across all the monitors in an area, meaning that attainment will now be determined by concentrations measures at each area's single worst-case monitor. This means that one can fully expect that attainment of the $PM_{2.5}$ NAAQS will be determined by the future near-road monitors. Given that it is highly likely that many areas' near-road monitors will have higher $PM_{2.5}$ readings than community-oriented monitors, the proposed NAAQS will become much more stringent and therefore more difficult to attain than it would be if based on the current monitoring network.

EPA has not attempted to characterize how much *more* stringent the standard will become as a result of these two changes to the NAAQS combined. A simple numerical example will have to illustrate how much it might impact the Proposed Rule's costs. Assume that the annual $PM_{2.5}$ standard is set at 13 $\mu g/m^3$, and a city's spatially-averaged annual $PM_{2.5}$ level based on the existing network of community-oriented monitors is 12.5 $\mu g/m^3$. It is in attainment and would face no cost from the Proposed Rule. However,

[1] See, for example, pp. 3-162 to 3-164 of EPA's *Integrated Science Assessment* for PM.

if it starts up a near-road monitor and levels of $PM_{2.5}$ at that monitor are 20% higher than at the average community-oriented monitor,[2] suddenly the $PM_{2.5}$ level determining that city's attainment status will rise to 15 $\mu g/m^3$. The city will be thrown into nonattainment, and probably face high costs of attainment, given the degree of additional ambient concentration reduction that it will need to accomplish. Thus, without the monitoring change the standard imposes no cost, but with the monitoring change it imposes large costs. Note that attainment costs increase while the NAAQS level itself is kept the same.[3] There is no evidence that EPA's analysis has accounted for how these elements of its Proposed Rule are likely to affect attainment status, and EPA's cost estimate is therefore likely underestimated.

<u>Section 2</u>. The Secondary Standard for Visibility Is Complex, Highly Sensitive, and Implies Different Regions Will Face Very Different $PM_{2.5}$ Limits

The proposed secondary standard for urban visibility is 28 to 30 deciviews, based on the 90^{th} percentile of 24-hour average $PM_{2.5}$ measurements (over a 3-year period). A deciview (dv) is very complex, and only readily understood by people who have spent a good deal of time working with visibility metrics. The complexity of this metric is illustrated by simply writing down some of the formulas by which it is calculated from monitored $PM_{2.5}$ concentrations:[4]

---

[2] Near-road monitoring was introduced with the NO_2 NAAQS. The NO_2 *Integrated Science Assessment* finds that NO_2 near roads can be twice the levels at other monitor locations. Spatial gradients of $PM_{2.5}$ are probably quite different from those of NO_2 away from roadways, but this and information in the $PM_{2.5}$ *ISA* (pp. 162-164) indicates that a 20% increment such as is used in this example is probably not unreasonable.

[3] The health risks EPA attributes to $PM_{2.5}$ do not change just because of the change in how attainment is determined – that is still calculated based on composite monitor levels, because those are the basis for the epidemiological associations.

[4] These formulas are in the Proposed Rule at p. 257 and p. 358, respectively.

$$\text{Deciview (dv)} = 10 \ln (\textit{bext} / 10\ \text{Mm}^{-1}),$$

where the value of *bext* is calculated from estimates of the concentrations of each of the individual $PM_{2.5}$ components listed in the next equation:

$$\begin{aligned} \text{PM2.5 } \textit{bext} = \; & 3 \times f(\text{RH}) \times [\text{Sulfate}] \\ & + 3 \times f(\text{RH}) \times [\text{Nitrate}] \\ & + 4 \times [\text{Organic Mass}] \\ & + 10 \times [\text{Elemental Carbon}] \\ & + 1 \times [\text{Fine Soil}] \end{aligned}$$

The final term in this complex calculation is *f*(RH), which is itself a complex and non-linear function of the location's relative humidity, RH. Thus, there many steps to calculate a deciview, precluding any rule of thumb or intuitive approximation of what a deciview means in terms of a limit on $PM_{2.5}$ concentrations.

The formulas are provided only to illustrate the complexity, and not because many readers will gain much understanding from reading them. The important points that they can be used to demonstrate, however, are: (a) seemingly small changes in the deciview level imply much larger changes in ambient $PM_{2.5}$ concentrations; and (b) a uniform national deciview standard will limit ambient $PM_{2.5}$ concentrations to very different levels in different cities.

One deciview is supposed to represent the amount of change in visibility that a normal eye can discern under optimal viewing conditions. It is a logarithmic function of visibility (called "light extinction," which is the term labeled "b_{ext}" in the above formulas) because people's ability to perceive visibility changes is logarithmic, meaning that approximately a 10% change in light extinction must occur before people can detect it.

Thus, a range from 28 dv to 30 dv (seemingly a 7% difference between the two levels) implies roughly a 20% range in visibility. For a given location with a fixed $PM_{2.5}$ constituent mix and relative humidity, the 2 dv range under consideration also implies about a 20% range in the implied $PM_{2.5}$ limit. This is quite a wide range from the perspective of the stringency of attaining such limits.

For example, in a city similar to Washington DC during a period of moderate humidity (*e.g.* about 70% RH), the 2 dv range implies a $PM_{2.5}$ concentration range from about 23 $\mu g/m^3$ to 28 $\mu g/m^3$.[5] This is a quite wide range of stringency – and also quite stringent given that it applies to 24-hour average $PM_{2.5}$.[6] Notably, EPA requests comment on a deciview standard as low as 25 dv. For a location with the above illustrative mix of $PM_{2.5}$ and relative humidity, this yet-lower visibility standard would imply a 24-hour average $PM_{2.5}$ concentration of about 17 $\mu g/m^3$! In short, s*eemingly small changes in the deciview level imply much larger changes in ambient $PM_{2.5}$ concentrations.*

EPA proposes a uniform national deciview standard, but this would imply a standard that is non-uniform across the country in terms of the limits it would imply on $PM_{2.5}$ concentrations. For example, consider what a uniform visibility standard of 28 dv would mean for different areas. For cities with the same $PM_{2.5}$ mix as in the prior example (*i.e.*, similar to the Washington DC mix), 28 dv would imply a $PM_{2.5}$ limit

[5] These calculations assume the $PM_{2.5}$ is 50% sulfate, 10% nitrate, 25% organic mass, 10% elemental carbon and 5% fine soil.

[6] For purposes of comparison, the 24-hour average primary standard is 35 $\mu g/m^3$.

ranging anywhere from 18 μg/m^3 in areas with very high relative humidity[7] to 43 μg/m^3 in areas with very low relative humidity. Similarly, if the relative humidity is held at about 50%, but the $PM_{2.5}$ mix is altered from sulfate-dominated to much less sulfate-dominated (more like Western areas of the U.S.)[8], the implied $PM_{2.5}$ limits for the same 28 dv standard would vary from about 29 μg/m^3 to 33 μg/m^3 for those two types of areas. In short, *a uniform national deciview standard will limit ambient $PM_{2.5}$ concentrations to very different levels in different cities.*

The visibility standard's difficulty to attain is also very sensitive to the form of the standard, such as its percentile and its averaging period. The numerical examples above suggest that depending on local conditions, the $PM_{2.5}$-equivalent limit of 28 dv may vary from 18 μg/m^3 to 43 μg/m^3 on a 24-hour average. It is not possible without further study to determine how much lower than the primary daily standard of 35 μg/m^3 (24-hour average) a location's 28-dv $PM_{2.5}$-equivalent limits must be before the visibility standard would actually be more stringent than the health-based $PM_{2.5}$ NAAQS. That is because even though both are based on 24-hour average data, the health-based standard is based on the 98th percentile value while the visibility standard is proposed to be based on the 90th percentile value. Differences in variability in relative humidity among locations make it difficult to expect it to be possible to develop even a rule of thumb for a relationship between the two percentiles.

---

[7] But less than the 90% RH screen provided for in the Proposed Rule.

[8] Assuming for this example that the less sulfate-dominated mix is 20% sulfate, 15% nitrate, 30% organic mass, 15% elemental carbon and 20% fine soil.

It is also important to note the sensitivity to the averaging period. In drafts of EPA's *Policy Assessment for the Review of the PM NAAQS,* EPA considered a 1-hour averaging period and a range of other percentiles. Based on analyses presented by EPA, I estimated that the 30 dv standard based instead on a *maximum daily 1-hour averaging period* (90th percentile) would have been more stringent than the primary daily standard (98th percentile) in 12 of 15 cities EPA had analyzed – in most cases being as stringent as a daily primary standard lower than 25 µg/m^3 (compared to the actual primary standard of 35 µg/m^3).[9] I also estimated that a shift of the form from 90th percentile to *98th percentile* could halve the equivalent $PM_{2.5}$ limit stated in the same form as the primary standard. Thus, what may seem to be minor changes in the proposed deciview level of the standard, such as a different averaging period or a different percentile, can make vast differences in its stringency. Shorter averaging periods or higher percentile requirements could easily render a visibility standard of 28 to 30 dv much more stringent than the primary standards for most areas of the U.S.

<u>Section 3</u>. EPA's Analysis Methods Do Not Credibly Identify a Visibility Level that Adversely Affects the Public Welfare

A secondary standard for urban visibility requires reasonable determination regarding what urban visual air quality (VAQ) levels are harmful to the public welfare. To make its determination in the Proposed Rule, EPA is relying a type of public survey it calls the "VAQ preference study" method. This is a highly simplistic survey, in which individuals are shown photographs of the same vista under a range of different visibility

[9] Anne E. Smith, *Technical Comments on Chapter 4 of EPA's "Policy Assessment for the Review of the Particulate Matter National Ambient Air Quality Standards" (First External Review Draft)*, prepared at request of Utility Air Regulatory Group and submitted to PM Docket with UARG's Comments, April 26, 2010, Table A-1.

conditions, and asked to rate whether the VAQ in each photograph is "acceptable" or "unacceptable." The VAQ at which EPA considers public welfare to be adversely affected (the "VAQ cutpoint") is the VAQ level that at least 50% of survey respondents deem unacceptable.

No theoretical formulations exist that justify this method as a measurement of public welfare effect. It is, however, at odds with much of the literature on the psychology of survey design and preference elicitation. Only four such surveys have been performed since the first such survey was conducted twenty years ago, each in a different location and each producing statistically significantly different estimates of a VAQ cutpoint. Given its central role in setting a NAAQS with potentially significant compliance challenges, the scientific validity of this method is an important question.

Earlier in the current PM NAAQS review cycle, I led a project that investigated the ability of this method to identify a robust estimate of the VAQ cutpoint that would not be sensitive to slight changes in the questionnaire used. We performed a controlled experiment with the survey instrument that to my knowledge is the the only known exploration of the scientific validity of the VAQ preference study method. We replicated the survey instrument from one of the existing VAQ preference studies. Then we ran two variants of the same survey in which the only change was to show respondents a different range of VAQ levels. We found that merely showing a different range of VAQ levels generated very different estimates of the VAQ cutpoint – differences that were also statistically significant. Other questions asked during these surveys indicated that respondents in each survey variant were equally well able to relate the particular range of

VAQ shown to them in the photographs to the full range of visibility conditions they experience in daily life. Thus, the significantly different responses across survey variants regarding what VAQ levels are "acceptable" cannot be attributed to inability on the part of respondents to discern when they were not being shown the entire range of actual visibility conditions.

This study provides strong evidence that the VAQ preference survey method that EPA is relying on does not actually estimate individuals' absolute preferences regarding VAQ, because an absolute preference structure would not be malleable to the particular levels of VAQ over which their preferences are elicited. Although this study was documented and provided to EPA in formal comments,[10] EPA has elected to ignore its clear and fundamental implication that EPA is not using a credible method for determining if urban visibility degradation is adversely affecting the public welfare under the current PM NAAQS. EPA's method also cannot credibly identify a level at which such adverse effects would occur.

Section 4. EPA's Acknowledgement of Uncertainty in Risks from $PM_{2.5}$ At Low Levels Is Not Reflected as Uncertainty in its Estimates of Benefits of the Proposed PM NAAQS

EPA's estimates of the health benefits of the proposed PM NAAQS in its RIA are far more uncertain than EPA admits. The Administrator's own rationale for setting the primary annual standard at a level no lower than 12 $\mu g/m^3$ is based on the greatly

[10] Anne E. Smith and Sabrina Howell, *An Assessment of the Robustness of Visual Air Quality Preference Study Results*. CRA International. Washington, DC. March 30, 2009. http://yosemite.epa.gov/sab/sabproduct.nsf/B55911DF9796E5E385257592006FB737/$File/CRA+VAQ+Pref+Robustness+Study+3+30+09+final.pdf.

increasing uncertainty that the Administrator expresses regarding EPA's own projections of health benefits from yet-lower $PM_{2.5}$ levels. For example, the Proposed Rule states:

> Based on consideration of the composite monitor annual mean $PM_{2.5}$ concentrations involved in estimating long-term exposure-related mortality, the Risk Assessment has higher confidence in using those concentrations that generally fall well within the range of ambient $PM_{2.5}$ concentrations considered in fitting the concentration-response functions used (*i.e.*, within one standard deviation of the mean $PM_{2.5}$ concentration reported in Krewski *et al.* (2009) for 1999-2000) as inputs to the risk model….With lower alternative annual standard levels of 12 μg/m^3 and 10 μg/m^3, the composite monitor annual mean values ranged from approximately 9.0 to 11.4 μg/m^3 and 7.6 and 8.9 μg/m^3, respectively. These concentrations are towards the lower end of the range of ACS data (in some cases approaching the lowest measured level) used in fitting the concentration-response functions, particularly for an annual standard level of 10 μg/m^3, and, thus, the Policy Assessment concludes there is less confidence in the risk estimates associated with these levels [10 and 12 μg/m^3] compared with those for the higher alternative annual standard levels considered.[11]

Building on this notion of increasing uncertainty about health risks at increasingly lower $PM_{2.5}$ levels, the Proposed Rule describes the Administrator's rationale for proposing an annual standard in the range of 12 μg/m^3 to 13 μg/m^3 in terms of such decreasing confidence. Some excerpts of the rationale are quoted below:

> In reaching decisions on alternative standard levels to propose, the Administrator judges that it is most appropriate to examine where the evidence of associations observed in the epidemiological studies is strongest and, conversely, where she has appreciably less confidence in the associations observed in the epidemiological studies**.** …
>
> She also recognizes …that there is significantly greater confidence in observed associations over certain parts of the air quality distributions in the studies, and conversely, that there is significantly diminished confidence in ascribing effects to concentrations toward the lower part of the distributions. …
>
> Based on the above considerations, the Administrator concludes that it is appropriate to propose to set a level for the primary annual PM2.5 standard within the range of 12 to 13 μg/m^3. . .While the Administrator recognizes that CASAC

[11] PM NAAQS Proposed Rule, p. 89 (emphasis added).

> advised, and the Policy Assessment concluded, that the available scientific information provides support for considering a range that extended down to 11 $\mu g/m^3$**,** she concludes that proposing such an extended range would reflect a public health policy approach that places more weight on relatively limited evidence and more uncertain information and analyses than she considers appropriate at this time.[12]

None of these admissions of uncertainty in the continuation of the risk association at ever lower $PM_{2.5}$ levels is reflected in EPA's estimates of the benefits of the Proposed Rule. This can be ascertained by the fact that the reported "range" of benefits associated with alternative levels of the proposed primary annual standard does not expand as that level is reduced. For example the range from the low to high end of the estimated benefits for the 13 $\mu g/m^3$ alternative standard (*i.e.,* \$88 million to \$220 million) is a factor of about 2.5. If EPA's benefits calculations reflected a growing uncertainty as $PM_{2.5}$ is reduced even further, then the range from the low to high end of the estimated benefits for the 12 $\mu g/m^3$ alternative standard would be greater than a factor of 2.5. But that range also is about a factor of 2.5 (*i.e.,* \$2.3 billion to \$5.9 billion). In sum, no expansion in uncertainty is reflected in EPA's benefits analysis as the alternative standard is tightened to the lowest level that the Administrator has confidence may be necessary to protect the public health with an adequate margin of safety.[13]

---

[12] PM NAAQS Proposed Rule, pp. 154-163 (emphasis added).

[13] In fact, experience from past EPA RIAs tells us that the range it states for its benefits estimate is defined by deterministic application of two different relative risk estimates from the epidemiological models, both of which are given full consideration in the Administrator's review of the weight of evidence that leads her to conclude that the lowest level of the standard that she has confidence in considering is 12 $\mu g/m^3$. Specifically, they are the 2002 study by Pope *et al.* of the American Cancer Society cohort at the lower bound and the 2006 study by Laden *et al.* of the Harvard Six-Cities cohort at the upper bound. The RIA does not treat either study's risk associations as any less certain at the lowest $PM_{2.5}$ levels than at the highest levels.

Thus, while uncertainty prevents the Administrator from being able to justify setting the standard lower, EPA's benefits calculations still take full credit for predicted health risk reductions that come from changes in $PM_{2.5}$ that occur at concentrations already well below 12 $\mu g/m^3$. The benefits analysis is thus inconsistent with the rationale for the standard, and projects an overstated degree of certainty.

Section 5. Doubts that the Chronic Mortality Risk Associations Are Causal Continue to Emerge

Statistical studies alone cannot provide strong evidence of causality even when many such papers have been published. This is because of the possibility of *systematic biases*, which can cause all of the statistical results from multiple different studies to be wrong for the same reason. The potential for systematic bias should not be ignored in deciding whether this body of purely statistical evidence is detecting an association that is causal.

Systematic biases can occur if the statistical studies have relied on similar methodologies and similar data sources. This is certainly the situation for the chronic mortality risk epidemiological literature that EPA is relying on for over 90% of its benefits estimate. All of the chronic mortality risks studies that EPA relies on to draw its conclusion of causality draw from the same fundamental universe of data, because they all sample individuals across the U.S. and assess the correlation between their local monitors' $PM_{2.5}$ levels and their relative mortality risks after attempting to control for the broad swath of much stronger determinants of risk (*e.g.*, age, sex, diet, smoking habits, and socioeconomic factors). Controlling effectively for these other factors is the key to getting a sound answer, yet *all* of the studies are reduced to using the same approximate data and feature high rates of error in assigning those variables to individual cohort members. In any single study, there is a good chance that the controls for the primary determinants of mortality risk are incomplete, and some confounding remains to bias the association estimated for $PM_{2.5}$. Unfortunately, all of these studies face the same problem, in a systematic way, because they all rely on the same types of data and face the same fundamental data limitations.

An innovative approach to exploring chronic $PM_{2.5}$ mortality risk described in a 2011 paper in the *Journal of the American Statistical Association*[14] finds that confounding could be playing a significant role in the statistical findings of positive $PM_{2.5}$-mortality associations. It does so by developing a mathematical method for estimating two separate risk coefficients, or "beta" values: one that is based on the changes in average $PM_{2.5}$ over time that is shared across the nation, and the other that is based on the changes in $PM_{2.5}$ that occur only within a city. If $PM_{2.5}$ has a causal relationship with mortality risk, for any effect that is detected for a $PM_{2.5}$ reduction that occurs nationally, a comparably sized effect should also be detected for $PM_{2.5}$ reductions that occur only locally. The authors state that "absent confounding or other model misspecification, the two estimates should be similar."[15]

Their analysis finds a relative risk from $PM_{2.5}$ just as other chronic studies have. However, they find that relative risk is entirely due to the beta for the national $PM_{2.5}$ trend, while declining $PM_{2.5}$ appears to contribute essentially zero risk reduction when it

[14] Sonja Greven, Francesca Dominici, and Scott Zeger, "An Approach to the Estimation of Chronic Air Pollution Effects Using Spatio-Temporal Information," *Journal of the American Statistical Association*, Vol. 106(494): 396-406, June 2011.
[15] *Ibid*, p. 397.

occurs within a city at a rate that differs from the nationwide trend. The authors express concern that they have found evidence of confounding in the $PM_{2.5}$ chronic risk association.

Since their methods effectively reproduce the type of relative risk that other chronic risk papers find when they also estimate only a single risk coefficient, this new paper offers highly suggestive evidence that all of the large body of statistical studies of mortality risk from chronic $PM_{2.5}$ exposure may – systematically – be detecting a non-causal association. It is unimportant whether those observed associations are confounded due to a missing explanatory variable for another pollutant, socioeconomic factor, or locational factor (such as noise from traffic) – a conclusion that the association is non-causality means that mortality risk will still not respond to changes in $PM_{2.5}$. It means that risk analyses based on those studies' results will all predict benefits where there will actually be none. Thus, confidence in EPA's benefits estimates, which are predicated on the *presumption* of causality, is overstated as long as the question of non-causality in the chronic mortality associations remains unresolved.

EPA's science assessment, which is the source of EPA's assertion that the chronic mortality risk associations can be viewed as causal, was written before the 2011 paper was published. When strong contradictory evidence emerges, as is the case now, it would seem prudent public policy to more closely examine and deliberate the new evidence before locking into major and costly new regulatory actions.