From: Joel Schwartz, American Enterprise Institute

To: Richard Bode, Linda Smith and Hien Tran, California Air Resources Board

Re: CARB's update of methodology to estimate premature deaths from fine particulate matter ($PM_{2.5}$)

Date: August 29, 2006

Via E-Mail: rbode@arb.ca.gov; lsmith@arb.ca.gov; htran@arb.ca.gov

---

I'm writing to provide comments on CARB's efforts to update its methodology for estimating premature deaths associated with $PM_{2.5}$ exposure. These comments provide more detail on the oral comments I delivered at CARB's August 21, 2006 workshop. I've also attached three of my papers that elaborate on some of the points I make in this letter.

The discussions and handouts at the August 21 workshop indicate that CARB's approach to evaluating the association of $PM_{2.5}$ and mortality tends to omit contrary evidence and to uncritically accept supportive evidence. This would cause CARB to overstate the magnitude and certainty of the association of air pollution and premature mortality.

In addition, CARB's plan for peer review of its studies ensures that CARB's report will be reviewed by scientists who already believe in the validity of the methods and results of air pollution epidemiology studies. EPA's Expert Elicitation (EE) likewise suffers from a bias toward scientists who are already supportive of EPA's and CARB's views and regulatory goals. Thus, using EPA's EE as part of CARB's review process will further bias CARB's review toward exaggeration of harm from $PM_{2.5}$.

In order to ensure that CARB's analysis reflects the real-world validity of $PM_{2.5}$ studies, and the real-world likelihood of harm from current, historically low levels of $PM_{2.5}$, I offer the following recommendations:

1. CARB should step back and assess whether observational epidemiology studies are capable of providing reliable information on the existence of small risks. Observational studies are the main justification for the claim of a causal association between air pollution and premature death, but they are also the weakest form evidence.

   As I summarize briefly below and discuss in more detail in the attached documents, there is good reason to discount the results observational studies, due to the inherent weaknesses and biases in the methods themselves, and due to the clear influence of data mining and publication bias. These factor all work to inflate apparent harm from

air pollution. In addition, experimental studies with both humans and animals don't support a causal air pollution-mortality association, contradicting the observational studies.

There are thousands of observational studies claiming to provide support for a causal association between low-level air pollution and risk of death. But implementing invalid techniques over and over again doesn't improve their validity.

2. CARB should be careful not to omit contrary evidence from its analysis and should take a more critical look at studies claiming to support a causal association between air pollution and mortality. For example, at the August 21 workshop, whenever CARB staff referred to specific research results, only results congenial to CARB's interests and regulatory goals were mentioned. I provide more examples of this type of selection bias in the attached papers.

3. To the extent that CARB wishes to include EPA's Expert Elicitation (EE) as part of its evaluation, the EE should be understood to represent the opinions of researchers who are already supportive of the epidemiological methods and results, and of EPA and CARB's views.

   EPA selected the experts based on mainly on how many air pollution epidemiology papers they've published. But the scientists who choose to do mainly air pollution epidemiology for their careers are clearly going to be scientists who already believe in the validity of the methods. Epidemiologists who believe observational epidemiology is not capable of providing reliable information on the existence of small risks are unlikely to devote their careers to performing such studies.

   Thus, selecting only scientists with a large number of air pollution epidemiology publications means that $PM_{2.5}$ health effects will be evaluated by a narrow group of believers, rather than by the wider community of experts in the mathematical techniques and research methods used in air pollution epidemiology (e.g., epidemiologists in other health fields, statisticians, and econometricians). This wider community of experts might not have as much faith in the validity of the methods as the narrow group of researchers who have chosen to work in air pollution epidemiology.

   A further source of bias is that regulatory agencies fund much of the research in air pollution epidemiology. These agencies decide what questions are asked and who is funded to answer them. This virtually guarantees that among the community of air pollution epidemiologists, which is already a self-selected group of believers in low-level air pollution risks, the researchers with the most publications will also be the ones whose views are most closely aligned with the views and agendas of regulatory agencies and environmental groups.

   Because of these inherent biases, it is not appropriate for CARB to consider EPA's EE as a valid analysis of whether or to what extent low-level air pollution causes premature death.

4. CARB's own analysis is at risk of suffering from similar selection biases. To ensure that CARB's report receives a genuine critical evaluation before its release, CARB should include among its advisors and peer reviewers (1) epidemiologists who have

provided evidence against the existence of a causal association between air pollution and mortality and/or against the validity of the epidemiological methods, and (2) researchers who are not air pollution epidemiologists, but who are expert in implementing and elaborating the mathematical techniques used in air pollution epidemiology, and of assessing the real-world validity of causal inferences based on those techniques, including econometricians, statisticians, and researchers in other sub-fields of risk factor epidemiology.

5. Putting aside the fundamental concerns about whether estimates based on observational studies represent real risks, it is important to find out why different researchers come up with such different results for $PM_{2.5}$ effects.

   For example, Pope et al. (1995 and 2002) analyzed the American Cancer Society's (ACS) CPS II cohort and reported a nationwide average relative risk (RR) of 1.069 per 10 $\mu g/m^3$ $PM_{2.5}$ for the 1982-89 follow-up period, and 1.04 for the 1982-98 follow-up period [1][2]. However, the Health Effects Institutes's (HEI) reanalysis of the 1995 study reported RRs below 1.0 for California specifically (see Figure 21, p. 197) [3].

   Jerrett et al. (2005) also analyzed the CPS II cohort [4]. Focusing on Los Angeles and using techniques developed specifically for their analysis, they reported RRs that are 2 to 3 times greater than the national-average RR reported in Pope et al. (2002).[a]

   Enstrom (2005) analyzed the ACS CPS I cohort data [5]. For an initial follow-up period of 1973-82, he reported RRs of about 1.04. But the RRs dropped to 1.0 or less from 1983 onward.

   These studies used different modeling approaches and controls for confounding, so it is difficult to determine the extent to which any of them reflect a real-world causal link between $PM_{2.5}$ and mortality.

   The lower RR in Pope et al. (2002) when compared with Pope et al. (1995) also suggests that whatever the $PM_{2.5}$ effect size, it is dropping with time. Based on data published in the two Pope et al. studies, it can be roughly calculated that the $PM_{2.5}$ RR declined from about 1.069 for 1982-89 to about 1.019 for the 1990-98 follow-up period. Based on the confidence intervals in the Pope et al. papers, the 1990-98 RR would be statistically insignificant. Enstrom (2005) summarized similar temporal declines in RRs reported for other cohorts, including the CPS I, Six Cities, and Veterans.

   To find out what is causing all of these different results, CARB should commission reanalyses to confirm that the original results can be replicated and to determine how robust and reliable the various results are to different specifications and approaches.

   Ideally, three separate researchers should perform these reanalyses: one should be a proponent of an air pollution mortality link, such as C. Arden Pope or Joel Schwartz, one should be a critic of the link, such as Suresh Moolgavkar or Fred Lipfert, and one

[a] Although it should be noted that the RR for $PM_{2.5}$ was statistically insignificant in models with the most extensive adjustments for confounding.

should be a relatively "moderate" researcher, if such can be found. Perhaps Sverre Vedal or Richard L. Smith would be good candidates. These analyses would ideally also be reviewed by other air pollution epidemiologists, as well as by other relevant experts, such as econometricians and statisticians.

Including researchers with different points of view will ensure vigorous testing, critique, and review of the validity of any given approach to analyzing the data and of the conclusions each researcher draws from the various analyses.

In the remainder of this memo I elaborate on a number of the issues raised in the recommendations above. Once again, please refer to the attached papers for additional details.

Observational Epidemiology Is Probably Not Capable of Providing Reliable Evidence on the Existence of Small Risks

The claim that air pollution at current U.S. levels is killing people rests almost solely on results of observational studies. But a number of researchers have provided evidence that observational studies are simply not capable of providing reliable information on the existence of small risks, such as those claimed for air pollution [6-8].

Another source of evidence against the reliability and validity of observational studies is the number of cases in which conventional medical wisdom justified by observational studies has been overturned or drastically scaled back by results from randomized trials. Among the steadily increasing list of examples are the effects of a low-fat diet, hormone-replacement therapy, Vitamin A supplementation, and calcium supplementation. Randomized trials showed that hormone-replacement therapy (HRT), a low-fat diet, and Vitamin A supplementation didn't reduce heart disease risk, as had been thought based on observational results. In fact, randomized trials showed that Vitamin A supplements increased the risk of cardiovascular mortality and lung cancer. A low-fat diet also didn't reduce colorectal cancer risk in women, and if it reduces breast cancer risk, the benefit is far tinier than researchers expected based on observational results.

Randomized trials continue to demonstrate that observational studies are often not reliable. The observational studies are based on the assumption that once you've controlled for known confounders, any residual correlation between a health outcome and a risk factor represents a causal connection. Experience shows that this assumption is often false. You can't make non-experimental data look like randomized controlled data just by controlling for a few imperfectly measured confounding factors.

In fact, the situation is even worse in air pollution epidemiology, because the putative risks are so much smaller than in the medical studies mentioned above. For example, based on observational studies, researchers believed that not being on HRT increased women's cardiovascular risk by a factor of 2 (i.e., 100%). But even this large apparent risk turned out to be spurious when randomized controlled trials genuinely removed bias from confounding factors.

In contrast, the putative risks of air pollution are on the order of a few tenths of a percent to several percent—much smaller than in the HRT case. Yet air pollution epidemiologists are claiming that they can reliably pick out these tiny putative risks from within a sea of

confounders that have much larger associations with mortality risk. Not only are the methods unreliable for this purpose, as I discuss in the next section, the unreliability is not random, but, due to publication bias and data mining, is biased in the direction of inflating the apparent risks of air pollution.

Research that has been ignored by regulators continues to demonstrate this problem of uncontrolled confounding. Among many examples, the Health Effects Institute reanalysis of the American Cancer Society study showed that adding migration rates into and out of cities into the statistical modeling caused the apparent harm from $PM_{2.5}$ to disappear [3]. Migration isn't the only confounding factor that was not properly accounted for. The $PM_{2.5}$ effect disappeared in a number of other sensitivity analyses as well. The HEI reanalysis also showed that $PM_{2.5}$ appeared to kill the moderately active, but not the sedentary or the very active. Once again, this biologically implausible result suggests residual confounding rather than real causal effects.

Keatinge and Donaldon (2006) have shown that time series air pollution studies have not properly controlled for modifiers of heat stress [9]. Time series studies do control for temperature. But at any given temperature, both lower winds and greater sunshine increase heat stress. Once they added in these effects, the apparent harm from ozone and PM_{10} disappeared.

You can never be sure you've properly controlled for all important confounders. The medical studies show that doing so is very difficult, even when the putative risks are much larger than those claimed for air pollution. Most importantly, as noted at the beginning of this section, the tools of observational epidemiology are probably inherently incapable of providing reliable information on the existence of small risks.

Publication Bias and Data Mining Exaggerate Apparent Air Pollution Health Effects

Researchers have shown that publication bias and data mining inflate apparent effects in many fields, not just air pollution epidemiology. But it is a particular problem in air pollution studies [11-14]. The data are easy to obtain so it is possible for large number of studies to be performed. The data are very noisy and the biologically correct models are unknown, creating the risk that researchers will "find" patterns in the data that are due to chance alone. Often the studies are funded and the researchers selected by agencies with a vested interest in finding harm from air pollution. And researchers are more likely to seek publication of, and editors are more likely to want to publish studies that find effects, rather than studies that don't.

Researchers have shown that publication bias inflates apparent pooled air pollution effects by as much as a factor of 3 [11, 13]. Accounting for data mining shows that air pollution has no effect or even an apparent "beneficial" effect in a large percentage of plausible models [12, 15].

CARB's analysis should take account of the risk-inflating effects of data mining and publication bias.

Lack of PM Effects in Toxicology Studies

A key reason to be skeptical about whether air pollution is killing people at current levels is that researchers have been unable to kill animals with air pollution, even at levels several times greater than ever occur in ambient air [16]. CARB's $PM_{2.5}$ analysis should take account of the lack of toxicological evidence for death from ambient $PM_{2.5}$, both for whole $PM_{2.5}$ and various $PM_{2.5}$ components.

When doing this analysis of PM toxicology, CARB should be sure to look beyond the abstract of Sun et al. (2005) mouse study, which claimed to prove a cause-and-effect association between current $PM_{2.5}$ levels and heart disease, especially along with a high-fat diet [17]. Both researchers and the media hailed this study as providing the proof, which had so far been lacking in animal studies, that air pollution is causing heart disease, and therefore premature death, in Americans.

As I show in [18], the researchers used mice with 14 times the cholesterol levels of normal mice, making the study irrelevant for real-world PM risks. In fact, the very reason for using such unrealistic mice is that even massive $PM_{2.5}$ doses don't cause heart disease in normal mice. I also show, contrary to the researchers' claims, that the $PM_{2.5}$ doses were unrealistically high.

A substantial fraction of $PM_{2.5}$ in California is ammonium nitrate. NRDC's representative at the August 21 workshop asserted that one of the surprises from CARB's goods movement analysis was that nitrate PM accounted for far more mortality than diesel PM (DPM). There's actually nothing surprising about this. CARB used the same concentration response function for premature mortality for all PM components. Since there's a lot more ammonium nitrate than DPM in the air, CARB's method created the appearance that nitrate PM kills more people than DPM.

While there's little toxicological evidence that low-level PM can kill, the evidence against nitrate or sulfate PM as a cause of death is especially strong. To the extent that CARB wants to attribute deaths to particular $PM_{2.5}$ sources and components, nitrates and sulfates are especially strong candidates for removal from CARB's death tallies.

Intervention Studies: Less than Meets the Eye

At the August 21 workshop, staff particularly focused on intervention studies as providing evidence that reducing air pollution reduces mortality. Studies specifically mentioned included one on ending the use of soft coal in Dublin [19], and the shutdown of a steel mill in Utah [20]. These studies were discussed as if their evidence was definitive and robust. However, both suffer from serious weaknesses. First, although they are "intervention" studies, they are still *observational* intervention studies, so all the caveats above about confounding and bias are still applicable. But there are additional specific reasons to be skeptical of these studies' conclusions.

Dublin Coal Study. The Dublin study clams to demonstrate that premature mortality decreased due to PM reductions that resulted from Dublin's ban on the burning of bituminous coal on September 1, 1990. However, there was a large spike in mortality in winter 1990, shortly before the ban went into effect. This could have created the appearance that the drop in mortality after the coal ban was due to the reduction in black smoke due to the coal ban, even if the coal ban had nothing to do with it.

The authors controlled for flu outbreaks in their model using an indicator variable, and one of the five flu outbreaks during the study period did occur during winter 1990. But it is clear from the data that the mortality spike is not mainly a result of the flu outbreak. First, the winter 1990 mortality spike occurred for all causes of death, whereas only cardiovascular and respiratory deaths were anomalously high during other flu outbreaks. Second, the winter 1990 mortality anomaly was much greater than during other flu outbreaks.[b]

It would be interesting to see how much the winter 1990 anomaly affects the overall conclusions. The authors of the original study did not do any sensitivity analyses and did not even remark on the existence of the winter 1990 anomaly in their paper.

Even if the Dublin results are taken to have found a causal relationship between lower air pollution and fewer deaths, it's not clear that it has any lessons for air pollution in California. First, the study was based on black smoke levels in Dublin, which went from winter-average levels of 85 $\mu g/m^3$ before the coal ban down to 22 $\mu g/m^3$ after. Annual-average levels went from 50 $\mu g/m^3$ down to 15 $\mu g/m^3$. These are for black smoke alone, rather than total $PM_{2.5}$, so total $PM_{2.5}$ levels would have been substantially higher. The study is thus based on much higher average $PM_{2.5}$ levels than the levels of the federal or California $PM_{2.5}$ standards, or even the higher levels that occur in parts of California.

In addition, the study used outdoor black smoke levels as the exposure variable. But the coal was being used for home space and water heating. Indoor PM exposures would therefore have been much higher than outdoor exposures, further increasing the exposure levels when compared with current U.S. standards.

Utah Steel Mill Study. The Utah study claimed that reductions in PM_{10} associated with the closure of a local steel mill caused a reduction in respiratory hospital admissions. However other researchers have presented data that suggest the study suffers from confounding by cyclic epidemics of respiratory syncytial virus [21, 22]. Furthermore, even if the PM_{10} association is taken as causal, there is evidence that only the heavy metals in the PM are causal [23]. Yet regulatory efforts regarding PM are based on the assumption that total $PM_{2.5}$ mass is what matters. To the extent that only heavy metals are important, the vast majority of $PM_{2.5}$ reductions will provide no premature mortality benefits.

These are just two studies among many. But they provide key examples of CARB's failure to look beyond the abstracts of studies that claim to be finding causal connections between air pollution and mortality.

[b] It is also worth noting that the authors didn't actually have any data on flu outbreaks in Ireland. Instead, they assumed that a flu outbreak was occurring in any 14-day moving window in which the national mortality rate due to influenza or pneumonia was above the 95^{th} percentile.

Considering the Full Weight and Strength of the Evidence

CARB and EPA tend to omit contrary evidence and ignore weaknesses in the studies that purport to demonstrate causal associations between air pollution and health. I give many examples in the three attached papers, and have summarized a few of them here.

To ensure a valid evaluation of $PM_{2.5}$ mortality effects, CARB must consider not only the headline conclusions of studies that claim to have uncovered causal associations. Rather, CARB must consider the full weight and strength of the evidence, including evidence against causal associations of air pollution and mortality, weaknesses in the studies that purport to demonstrate a causal connection, evidence on the fundamental validity of the methods used to make causal claims, and selection biases in who performs and reviews the research.

In order to perform a valid evaluation of whether and to what extent current $PM_{2.5}$ levels are causing premature mortality CARB should cast a wider net in terms of both what studies it considers, how it evaluates their validity, and which researchers it chooses to provide expert advice and peer review.

Thank you for the opportunity to provide comments on CARB's review of methodologies for assessing the relationship between particulate matter and mortality. If you have questions or would like additional information, I can be reached at 916-203-6309 or jschwartz@aei.org.

Enclosures

J. Schwartz, *Rethinking the California Air Resources Board's Ozone Standards* (Washington, DC: American Enterprise Institute, September 2005), http://www.aei.org/doclib/20050912_Schwartzwhitepaper.pdf.

J. Schwartz, *Air Pollution and Health: Do Popular Portrayals Reflect the Scientific Evidence?* (Washington, DC: American Enterprise Institute, May 2006), http://www.joelschwartz.com/pdfs/AirPoll_Health_EPO_0506.pdf.

J. Schwartz, *Comments on EPA's Proposed Rule, National Ambient Air Quality Standards for Particulate Matter* (Washington, DC: American Enterprise Institute, April 17, 2006), http://www.joelschwartz.com/pdfs/Schwartz_PM25_NAAQS_041706.pdf.

References Cited

1. Pope, C.A., 3rd, Burnett, R.T., Thun, M.J., et al. (2002). Lung Cancer, Cardiopulmonary Mortality, and Long-Term Exposure to Fine Particulate Air Pollution. Jama, 287, 1132-41.
2. Pope, C.A., 3rd, Thun, M.J., Namboodiri, M.M., et al. (1995). Particulate Air Pollution as a Predictor of Mortality in a Prospective Study of U.S. Adults. Am J Respir Crit Care Med, 151, 669-74.
3. Krewski, D., Burnett, R.T., Goldberg, M.S., et al. (2000). Reanalysis of the Harvard Six Cities Study and the American Cancer Society Study of Particulate Air Pollution and Mortality, July, Cambridge, MA. Health Effects Institute.
4. Jerrett, M., Burnett, R.T., Ma, R., et al. (2005). Spatial Analysis of Air Pollution and Mortality in Los Angeles. Epidemiology, 16, 727-36.
5. Enstrom, J.E. (2005). Fine Particulate Air Pollution and Total Mortality among Elderly Californians, 1973-2002. Inhal Toxicol, 17, 803-16.
6. Smith, G.D. (2001). Reflections on the Limitations to Epidemiology. J Clin Epidemiol, 54, 325-31.
7. Taubes, G. (1995). Epidemiology Faces Its Limits. Science, 269, 164-69.
8. Ioannidis, J.P. (2005). Why Most Published Research Findings Are False. PLoS Med, 2, e124.
9. Keatinge, W.R. & Donaldson, G.C. (2006). Heat Acclimatization and Sunshine Cause False Indications of Mortality Due to Ozone. Environ Res, 100, 387-93.
10. Moolgavkar, S.H. (2005). A Review and Critique of the EPA's Rationale for a Fine Particle Standard. Regulatory Toxicology and Pharmacology, 42, 123-144.
11. Anderson, H., Atkinson, R., Peacock, J., et al. (2004). Meta-Analysis of Time-Series Studies and Panel Studies of Particulate Matter (PM) and Ozone. World Health Organization.
12. Koop, G. & Tole, L. (2004). Measuring the Health Effects of Air Pollution: To What Extent Can We Really Say That People Are Dying from Bad Air? Journal of Environmental Economics and Management, 47, 30-54.
13. Bell, M.L., Dominici, F., and Samet, J.M. (2005). A Meta-Analysis of Time-Series Studies of Ozone and Mortality with Comparison to the National Morbidity, Mortality, and Air Pollution Study. Epidemiology, 16, 436-45.
14. Lumley, T. & Sheppard, L. (2003). Time Series Analyses of Air Pollution and Health: Straining at Gnats and Swallowing Camels? Epidemiology, 14, 13-4.
15. Ito, K. 2003. Assocations of Particulate Matter Components with Daily Mortality and Morbidity in Detroit. Revised Analyses of Time-Series Studies of Air Pollution and Health, (pp. Boston: Health Effects Institute.
16. Green, L.C. & Armstrong, S.R. (2003). Particulate Matter in Ambient Air and Mortality: Toxicologic Perspectives. Regul Toxicol Pharmacol, 38, 326-35.

17. Sun, Q., Wang, A., Jin, X., et al. (2005). Long-Term Air Pollution Exposure and Acceleration of Atherosclerosis and Vascular Inflammation in an Animal Model. JAMA, 294, 3003-10.

18. Schwartz, J. (2006). Air Pollution and Health: Do Popular Portrayals Reflect the Scientific Evidence? May, Washington, DC. American Enterprise Institute.

19. Clancy, L., Goodman, P., Sinclair, H., et al. (2002). Effect of Air-Pollution Control on Death Rates in Dublin, Ireland: An Intervention Study. Lancet, 360, 1210-4.

20. Pope, C.A., 3rd. (1989). Respiratory Disease Associated with Community Air Pollution and a Steel Mill, Utah Valley. Am J Public Health, 79, 623-8.

21. Lamm, S.H., Hall, T.A., Engel, A., et al. (1994). PM10 Particulates: Are They the Major Determinant of Pediatric Respiratory Admissions in Utah County, Utah (1985–1989) Annals of Occupational Hygiene, 38 (supplement 1), 969-972.

22. Lyon, J.L., Stoddard, G., Ferguson, D., et al. (1996). An Every Other Year Cyclic Epidemic of Infants Hospitalized with Respiratory Syncytial Virus. Pediatrics, 97, 152-3.

23. Molinelli, A.R., Madden, M.C., McGee, J.K., et al. (2002). Effect of Metal Removal on the Toxicity of Airborne Particulate Matter from the Utah Valley. Inhal Toxicol, 14, 1069-86.