Discussion for "Are fine particulates killing Californians?"

Michael E. Ginevan, PhD.
M.E. Ginevan and Associates. Silver Spring MD 20901

Abstract:
This paper discusses the presentations made in the session "Are fine particulates killing Californians?" The central conclusion is that fine particulates are, in fact, probably not killing Californians. Moreover, the California Air Resources Board's (CARB) process for setting diesel emission standards raises a number of important questions about the scientific basis of regulatory standard setting in general. These range from how or whether one can set regulatory standards by relying on one or a few observational epidemiology studies to how one should organize a science review process for the standards themselves.

Key Words: Regulatory standards, multiple modeling, multiple comparisons, observational studies, PM 2.5, data sharing.

1. Introduction

The four talks in this session by S. Stanley Young (Young and Xia, 2012), Fred Lipfert (Lipfert and Young, 2012), James E. Enstrom (Enstrom, 2012), and Robert F. Phelan (Phelan, 2012) deal with two related issues that have a very substantial statistical component. The first issue is how should a regulatory agency deal with the problem of identifying potential adverse health effects of environmental pollution, and the issue second is how should a regulatory agency promulgate regulations designed to mitigate any adverse health effects identified. The focus of these four talks deal with the California Air Resources Board's (CARB) efforts to identify adverse effects of fine particulate air pollution and to promulgate regulations on diesel truck emissions which would presumptively reduce the health effects identified. The general impression one gets from these talks is that there are substantial deficiencies in both the process used to identify adverse health effects and in the subsequent promulgation of regulations designed to mitigate any adverse health effects identified.

2. Dr. Young's Presentation

Dr. Young's talk, "Variable Importance in Environmental Studies" highlights the two major issues in identifying actual effects in observational studies, multiple testing and multiple modeling and suggests that partitioning data sets to see if effects are consistent across data sets and examining the relative importance of variables like those on which the CARB diesel regulations are based are useful tools for the evaluation of the robustness of model results. For the case of consistency across subsets of the data, one important idea is that particulate air pollution effects are not consistent across regions of the country. Since the western region, which includes California, shows no evidence of particulate mediated health effects, the scientific basis of the California diesel regulations seems questionable.

The other important point raised by Dr. Young is that PM 2.5 levels are not a very important determinant of human mortality when compared to major factors like income levels. One could argue that one would not expect PM 2.5 levels to be a major determinant human mortality, because the expected effects are subtle. However, one must have major determinants like income in any model of life expectancy, and if the model used fails to capture even a small amount of the effect of the major determinants, the residual lack of fit could easily manifest as a spurious PM 2.5 effect.

Dr. Young also points out that because of the flexibility of the model building process the process becomes something of a Rorschach Test; different analysts can come up with very different interpretations of the same data.

Overall, Dr. Young's presentation convincingly calls into question the robustness and reliability of complex, exploratory data analyses applied to observational data as a basis for the setting of environmental standards.

3. Dr. Enstrom's Presentation

Dr. Enstrom's talk provides a more detailed look at the process surrounding the CARB diesel standard. In his talk he reiterates the idea that PM 2.5 health effects have not been demonstrated in California. He also reiterates that PM 2.5 effects are small and are estimated against a background of confounding factors whose effects are much larger in magnitude than the presumptive effects of PM 2.5. Dr. Enstrom also criticizes the exposure estimates used in the CARB modeling effort, and makes the point that even if one could accurately characterize a person's PM 2.5 exposure, the composition, and thus presumably the toxicity of PM 2.5 can vary from locality to locality. That is, identical PM 2.5 exposures may have very different toxicities. He also, quite properly, criticizes the fact that many key PM 2.5 data sets have been effectively "secret" for many years. That is, despite the fact that these data have been the basis of health studies that, in turn, have been the basis for setting more stringent, and much more expensive, air quality standards, these data have not been generally available to the scientific community.

However, the most disturbing aspect of his presentation is the idea that the CARB proceedings were focused on supporting a pre-determined conclusion. Dr. Enstrom cites a 2008 CARB report that used an analysis of U.S. national mortality and PM 2.5 data to estimate premature deaths from PM 2.5 in California. This seems questionable because California and indeed the western U.S. do not show negative effects of PM 2.5 on human health. However the first CARB report was set aside because the academic credentials of the principal author turned out to be falsified. Dr. Enstrom then cites some of his own work (Enstrom, 2005) and an unpublished special analysis of California subjects from Krewski et al. (2009), which both show no evidence of PM 2.5 effects on human health.

Following these negative results CARB funded another study to reexamine the question of PM 2.5 and health effects in California (Jerrett et al. 2011). This study is the focus of Dr. Lipfert's presentation which is discussed below. The Jerrett study includes many models that do not support a positive association between PM 2.5 levels and does not

provide what I would call a firm basis for setting environmental standards, yet CARB declared the study positive and proceeded to finalize their diesel truck regulations.

4. Dr. Lipfert's Presentation

Dr. Lipfert's talk focuses on the Jerrett study mentioned above. There are a number of striking points made in this discussion. First, the study used four different estimates of PM 2.5 exposure derived from different estimation. procedures Two of these, Kriging and inverse distance weighting are spatial interpolators, which attempt to estimate PM levels at a given space-time point as a weighted average of other space-time points where measurements exist. The other two are land use regression (LUR) and a variant of LUR based on Bayesian maximum entropy estimation. The problem is that there are only 112 PM2.5 measurement sites in the entire state of California, so there is a dual problem of which estimation method is best, and a more serious issue of, given the paucity of data available to parameterize exposure models, are any of these estimates good enough to be useful?

A second issue is that the modeling is done with and without indicator variables for certain geographic areas within California. This might not be problematic if the indicators made no difference, but according to Dr. Lipfert's presentation, the largest effect estimates are shown in models which include the indicator.

The largest issue, according to Dr. Lipfert's talk, is that the Jerrett Report includes some 400 models for PM 2.5 health effects, and not all or even most show evidence of PM 2.5 health effects. There appears to be a consistent elevation of PM 2.5 effects for ischemic heart disease but a countervailing depression in all non-cardiorespiratory disease effects, such that there is no effect on all-cause mortality. One might argue that PM 2.5 (assuming that this is a homogeneous toxic exposure) would affect only certain diseases, but a reasonable expectation would be that some diseases would be significantly elevated, while the elevation across all diseases would be markedly less. However, this is not what the results shown by Dr. Lipfert suggest. Some diseases are elevated others are depressed and the net result is an almost perfect null result.

My overall impression is that the Jerrett study falls considerably short of providing a basis for a causal conclusion that PM2.5 is killing Californians.

5. Dr. Phalen's Presentation

Dr. Phalen's talk presents the perspective of an advisor to the EPA regulatory process. He served on the U.S. EPA Clean Air Scientific Advisory Committee from 2007-2011. He makes a number of points of direct relevance to the California process for promulgating diesel regulations for PM 2.5. First, he reiterates the idea that the chemical composition of PM 2.5 and thus its toxicity can vary markedly. Thus standards based on particle mass are not sensible. He also points out that national standards are a poor approach to regional problems, because, as noted by earlier speakers, not all regions of the country have the same air quality problems.

Some of the issues presented by Dr. Phalen are more subtle. One is that advisors are not to consider tradeoffs. That is, one may reduce exposure to an air pollutant but if this reduction elevates the cost of energy or other goods, it may cause adverse health effects in low income sub-populations, which are, as Dr. Young pointed out, especially vulnerable to adverse health effects to begin with. Dr. Phalen also points out that "Use of limited questions in managed order causes problems." This refers to the common practice of managing the deliberations of advisory committees through highly restrictive "charges" to the committee. If one is a regulator one can often influence the input of the advisory committee by simply moving difficult points outside the charge of the committee.

There is also the issue that pollutants are considered one at a time and the related problem of margin of safety for risks. For the "one at a time" case, the issue is that exposure to air pollutants always involves exposure to mixtures of toxicants and the concentrations of the constituents of the toxic mixture are correlated across time and space. Take automobiles as an example. Car exhaust contains fine particulates, but it also contains nitrogen oxides, carbon monoxide, and hydrocarbons, as well as other potentially toxic constituents. High traffic areas will have elevated levels of fine particulates but will also have elevated levels of all other automobile related constituents, and because of photochemical reactions in the atmosphere may also have elevated levels of other toxic materials such as ozone. The statistical problem is how can one analyze for health effects of one toxicant while taking the mixture into account? If one statistically controls for all other pollutants while analyzing for effects of particulates, it seems quite likely that effect estimates for particulates will be minimized, but if one does an analysis for particulate effects without controlling for other toxicant exposures it seems equally likely that effect estimates for particulates will be maximized.

The idea of an adequate margin of safety would lead one to lean toward analyses that overstate risks because it results in effect estimates that are higher and thus exposure standards that are lower and thus inherently "safer." If this approach is applied across a mixture the result is a set of standards that are very low for each toxicant. This may not be any safer than standards which are set for mixtures and which would likely allow higher levels of individual "toxicants" (which may or may not in fact pose significant health risks because all we have are mixture exposures), but the one-at-a-time standards will surely be more expensive to meet, which brings up the idea of economic consequences discussed earlier.

6. Conclusions

It seems quite clear that the CARB process for setting diesel emission standards for particulates is flawed and needs to be revisited. More importantly these talks suggest that there are fundamental flaws in the way science is used in the regulatory process. The first issue is that one or a few observational studies cannot be taken as clear evidence of adverse health effects. One might then suggest that one should rely more on laboratory studies of animals. This raises the dual problems that the complex exposures in the real

world are difficult, if not impossible, to mimic in a laboratory setting and, in any case, the extent to which responses in high dose animal exposures are predictive of low dose exposures in human populations is unclear and varies depending upon the toxic material involved in the exposure.

A general point, made by many authors (e.g. Young and Karr, 2011; Ioannidis, 2005) is that one or a few observational studies are not a basis for declaring that effects exist. Moreover, broadly focused analyses of a single data set as in Jerrett et al. (2011) are a particularly problematic basis for inferring causal relationships, let alone dose response functions that might be used in quantitative standard setting.

A best practice approach will likely involve a comprehensive consideration of all available information within some sort of structured (and in my view statistical) process. Two recent attempts in this direction can be found in Adami et al. (2011) and in the 2008 IOM report on improving presumptive disability decisions (IOM, 2008). I add that the latter document, which I would expect most statisticians to miss, contains a highly detailed discussion of a Bayesian statistical framework for combining information to both judge the causal plausibility of toxic exposure / human disease relationships and develop best estimates of dose-effect relationships.

One concern I would expect to be voiced is that by requiring more information and more rigorous analysis before any regulation could be promulgated one would place human lives at risk. That is, if we don't regulate we may be killing people. The problem is that if we go on with the status quo, we are promulgating regulations for exposures that may or may not pose risks to human health but will certainly impose costs and tradeoffs on the regulated community. These cost and tradeoffs may, in turn cause real health effects that are larger than the hypothetical (and perhaps illusory) adverse effects the regulation was promulgated to prevent.

A second point is that any dataset used in any analyses which serve as a basis for promulgating regulations must be made available to all interested parties. In addition, an inventory of all of the analyses performed on these data (not just those that are published) should also be publically available. Making data available is a minimum requirement but is not sufficient for judging the robustness or reliability of a particular conclusion. Science magazine recently had a special section on data replication and reproducibility (see Jasny et al., 2011 for a summary). While these discussions were primarily focused on things like fraud detection in academic science, the ideas discussed have relevance to judging the scientific basis of proposed regulations. Making the data and analytical pedigree of regulatory analyses widely available certainly would result in every costly regulatory decision being the subject of intense review and scrutiny, but if the scientific basis of a regulation that may trigger billions of dollars in compliance costs will not stand such scrutiny, should it be promulgated in the first place?

My last point is that statisticians should work to produce more robust tools to identify health effects in observational data. Dr. Young's presentation at this meeting is a case in point. In examining variable importance and multiple endpoints it provides insight as to

the actual effects, if any, demonstrated in a given analysis. The Observational Medical Outcomes Partnership (OMOP) program (OMOP, 2012) is a much more comprehensive effort. However there is still an enormous amount of work to do, and statistical tools for defining the scientific basis for regulatory standards setting may pose special challenges. One question that is likely to arise routinely is how much scientific evidence is enough? The answer is likely to be problem specific. It seems reasonable that the evidentiary requirements for extremely expensive regulations like the CARB diesel standards should be higher than the evidentiary standards for regulations which have minimal compliance costs. The important thing is that we need to develop more rigorous tools for judging the evidentiary basis of regulatory decisions because, as the CARB diesel example suggests, the status quo is not working very well.

7. Some thoughts on funding.

Drs. Young and Enstrom both pointed out that CARB spent significant amounts of money on studies designed to support their standard setting process (some $750,000 on the Jerrett report alone). Indeed all regulatory agencies have substantial budgets focused on research aimed at identifying potential health hazards in need of regulation or quantifying dose-effect relationships that can be used to support standard setting. This is as it should be, but when the industries that are the object of regulatory activities fund scientists to explore the same problems, the resulting papers are often dismissed as "Industry sponsored studies" as though they were in some sense improper. I would suggest that both the regulators and the regulated have an equally good argument for funding research, and that the source of funding should be irrelevant to evaluating the quality of a study. My reason for mentioning this is that many academic researchers are reluctant to pursue industry funding (particularly in controversial areas) because of a perception that industry sponsored research is, in some sense, inherently biased. Certainly no researcher should allow the source of his or her funding to affect how they design studies, analyze data, or report results, but this should be a question of individual integrity not funding source. For the record, no participant in this session received any industry or government support for preparing their presentations.

References.

Adami, H., C.L. Berry, C.B. Breckenridge, L.L. Smith, J.A. Swenberg, D.Trichopoulos, N.S. Weiss, and T.P. Pastoor. 2011. Toxicology and Epidemiology: Improving the Science with a Framework for Combining Toxicological and Epidemiological Evidence to Establish Causal Inference. Toxicological Sciences 122, 223–234.

Enstrom, J.E. 2012. *Are Fine Particulates Killing Californians?* Presentation at Session #546, "Are Particulates Are Fine Particulates Killing Californians?" 2012 Joint Statistical Meetings, August 1, 2012 San Diego, California

IOM (Institute of Medicine). 2008. Improving the Presumptive Disability Decision-Making Process for Veterans. National Academies Press. Washington, DC.

Ioannidis, J.P.A. 2005. Why Most Published Research Findings Are False. PLoS Med 2(8): e124. doi:10.1371/journal.pmed.0020124.

Jasny, B.R., G. Chin, L. Chong, and S. Vignieri. 2011. Again, and Again, and Again …. Science 334, 1225.

Lipfert F. and S.S. Young, 2012. *Critical Review of Spatiotemporal Analysis of Air Pollution and Mortality in California Based on the American Cancer Society Cohort: Final Report. Principal Investigator: Michael Jerrett, PhD.* Presentation at Session #546, "Are Particulates Are Fine Particulates Killing Californians?" 2012 Joint Statistical Meetings, August 1, 2012 San Diego, California

OMOP (Observational Medical Outcomes Partnership) 2012. http://omop.fnih.org/node/20.

Phalen, R.F. 2012. *Improving the scientific advice provided by the Clean Air Scientific Advisory PM Subcommitte.* Presentation at Session #546, "Are Particulates Are Fine Particulates Killing Californians?" 2012 Joint Statistical Meetings, August 1, 2012 San Diego, California

Young, S.S. and A. Karr. 2011. Deming, data and observational studies. Significance Volume 8 #3, 116-120.

Young S.S. and J. Q. Xia, 2012. *Variable Importance in Environmental Studies.* Presentation at Session #546, "Are Particulates Are Fine Particulates Killing Californians?" 2012 Joint Statistical Meetings, August 1, 2012 San Diego, California