FRANCESCA DOMINICI, PhD
Professor of Biostatistics
Associate Dean for Information Technology

HARVARD | SCHOOL OF PUBLIC HEALTH

November 28 2012
From: Francesca Dominici, Scott L Zeger, Holly Janes, Sonja Greven
To: Bryan Hubbell, Jason Sack, EPA

Reply to: ***additional information to EPA for these two studies***

1) Janes H, Dominici**,** Zeger SL (2007) Trends in Particulate Matter and Mortality in 113 U.S. Counties, 2000-2002: Evidence on the Long Term Effects of Air Pollution. *Epidemiology,* 18: 416-423 (with discussion)
2) Greven S, Dominici F, Zeger SL (2011): An Approach to the Estimation of Chronic Air Pollution Effects Using Spatio-Temporal Information. *Journal of the American Statistical Association,* 106(494): 396–406.

These two studies targeted a very specific question: whether there is an association between month-to-month variations in mortality rates and month to month variations in the average PM2.5 for the previous 12 months (**global effect**) in the Medicare population. The fact that this particular study does not find an association at this specific time scale of variation of exposure to air pollution and for the study population considered cannot lead to a conclusion that all the other epidemiological studies are confounded and that air pollution is not causally linked to mortality. In this document we summarize our rationale for making this statement.

In these studies we decompose the **global effects** into two parts: 1) the association between the national average trend (NAT) in the monthly PM2.5 levels averaged over the previous 12 months and the national average trend (NAT) in monthly mortality rates (**national effect**); and 2) the association between the deviation of the community-specific trend from the NAT of PM2.5 and the deviation of the community-specific trend from the (NAT) of mortality rates (**local effect**). We decompose the global effect into a national effect plus a local effect because we hypothesize that the national effect is more likely to be affected by unmeasured confounding than the local effect. We also argue that if there are large differences between the local and the national effects then the global effect should not be reported without a more in depth investigation of confounding. We also acknowledge in the papers that differences between the local and national effects might be due to measurement error and not necessarily unmeasured confounding.

Our results do not invalidate previous epidemiological studies: In Janes et al and Greven et al, both using Medicare data, we did not find evidence of a local effect and we instead found evidence of a national effect. Although these results call for additional investigation of why we found these differences between the local and the national effects, these results do not invalidate results of the other cohort and multi-site time series studies. We summarize below why:

Department of Biostatistics
655 Huntington Ave
Building II, Room 441
Boston, MA 02115
www.hsph.harvard.edu

TEL 617 432 4908
FAX 617 432 5619
fdominic@hsph.harvard.edu

1. We eliminate spatial information, purposely. Our studies, on purpose, discard all the spatial variation in air pollution and mortality that is available in the data. This is the main and only information used in the cohort studies. Also, our studies cannot invalidate the time series studies, because they also do not rely on the day-to-day information. Again, our studies only focus on associations at monthly temporal scales and include a fixed effect for county in the regression model which purposely eliminate the spatial information when estimating the global effect. At the other end, the cohort studies by Pope et al 2009 NEJM, rely almost entirely on spatial information and they have developed approaches to adjust for measured and unmeasured confounding. In Table 1 of Janes et al, the percentage of the variability in their $PM_{2.5}$ variable is broken down by space, time, and space by time components. We found that over 90% of the variance in this data is attributable to spatial variation (which we ignore), and roughly only 5% attributable to the space by time component (which is what we used in our papers). Thus, when one considers that this wealth of information is not accounted for in that study, it is not as surprising that we see vastly different estimates of the $PM_{2.5}$/mortality relationship than in other studies that do exploit that variability.

2. Both the local and the national effects can be affected by unmeasured confounding. We do hypothesize that the national effect is more likely to be affected by unmeasured confounding than the local effect. However the unmeasured confounding for the local scale cannot be excluded either. For example, the national association can be confounded by a national downward trend in smoking that causes a reduction in mortality. As PM2.5 and mortality are both trending downwards nationally, not having a good measure of the effect of smoking in the model would cause an upward bias in the national association. An example that would cause the local association to be biased upwards would be if communities that were very 'health conscious and environmentally friendly' and thus reduced their air pollution more than the average community would also improve on other health-related measures such as smoking and obesity, causing drops in mortality larger than in the average community, even without the effects of the reduce in air pollution. On the other hand, an example that would cause a downward bias in the local association would be local manufacturing activity. For example, suppose a new plant is built that employs 3,000 workers and has a positive influence on the economic productivity of its local area. The improved economy causes reduced mortality relative to the regional trend, but also increases pollution relative to the regional trend. Such a phenomena would shift the local association toward zero. As individual- and location-level information, in contrast to cohort studies, is very limited for the Medicare data, we cannot exclude any of these potential sources of confounding in our studies.

3. Another important driver of the potential differences between the local and the national effects could be measurement error. We expect that measurement error could affect both these associations but it is hard to anticipate which one is more biased. If the exposure is measured with error, as we know is the case, one of the two components might be more precise than the other. Because the national scale component is an average of more independent monitors across a broader geographic extent, its "measurement error" (including local geographic variations and pure device measurement error) might be expected to be smaller, relative to the size of the PM signal. The local component might have greater measurement error as a proportion of its signal, and this might bias its coefficients, probably toward the null. More work is needed in this area.

4. <u>There might not be an association at the time scale and lag investigated in our studies.</u> Further, it is an open question as to what lag the "monthly" $PM_{2.5}$ levels should be calculated at. In the Janes paper this is calculated as the 12 and 24-month averages of $PM_{2.5}$ up to and including the current month. It is then assumed that that calculated exposure is associated with the current month's observed deaths. It may be the case that the exposure period does not stretch that far back, and is in fact over a shorter window. In Janes et al, the authors acknowledge that this is a question that warrants further research (see also Schwartz et al EHP 2008). Thus, it is entirely possible that these papers are looking for an association at a timescale for which no association truly exists.

In summary, in Janes et al 2007 and Greven et al 2011, we examined in unbiased fashion, the epidemiologic evidence relevant to estimating the effect of chronic air pollution exposure on survival. We have published all of our findings without regard to the outcome. None of the investigators have a financial conflict of interest.

Francesca Dominici, PhD
Associate Dean of Information Technology
Professor of Biostatistics
Harvard School of Public Health, Boston, MA

Scott L. Zeger, PhD
Vice Provost for Research
The Johns Hopkins University, Baltimore, MD

Sonja Greven, PhD
Assistant Professor
Department of Statistics
Ludwig-Maximilians-Universität München
Ludwigstraße 33
Munich, Germany

Holly Janes, PhD
Affiliate Assistant Professor, Biostatistics
Fred Hutchinson Cancer Research Center
University of Washington, Seattle, WA