UNITED STATES ENVIRONMENTAL PROTECTION AGENCY
Office of Air Quality Planning and Standards
Research Triangle Park, NC 27711

April 7, 2011

MEMORANDUM

SUBJECT: $PM_{2.5}$ Distributional Statistical Analyses

FROM: Pradeep Rajan, OAR/OAQPS/HEID /s/
Mark Schmidt, OAR/OAQPS/AQAD /s/
Beth Hassett-Sipple, OAR/OAQPS/HEID /s/

TO: PM NAAQS Review Docket EPA-HQ-OAR-2007-0492

Overview

This memorandum documents the analyses of population-level data from epidemiological studies conducted by EPA staff for the review of the particulate matter (PM) national ambient air quality standards (NAAQS). The purpose of these analyses was to employ distributional statistics to inform staff conclusions in the *Policy Assessment* on a range of alternative standard levels that is appropriate to consider in identifying a primary annual $PM_{2.5}$ standard that provides appropriate protection for health effects associated with both long- and short-term $PM_{2.5}$ exposures.[1]

In reaching staff conclusions on the range of annual standard levels appropriate to consider, we expanded our general approach for identifying alternative standard levels beyond the consideration of a single statistical metric (i.e., mean) from an epidemiological study, to the extent such information was available. In so doing, we employed distributional statistics to identify the broader range of $PM_{2.5}$ concentrations that were most influential in generating health effect estimates in epidemiological studies as well as to characterize where the bulk of the data exist. Thus, we considered the range of $PM_{2.5}$ concentrations where the data (i.e., population-level) from epidemiological studies were most dense, and at what range below the long-term mean $PM_{2.5}$ concentrations where there is a comparative lack of data and our confidence in the associations is appreciably less. We recognize there is no one percentile value within a given distribution that is the most appropriate or "correct" way to characterize where our confidence in the associations becomes appreciably lower. Consequently, we have determined that the range from the 25^{th} to 10^{th} percentiles is a reasonable range to consider as a region where we have appreciably less confidence in the associations observed in the epidemiological studies.

[1] US EPA (2011). Policy Assessment for the Review of the Particulate Matter National Ambient Air Quality Standards. Office of Air Quality Planning and Standards, U.S. Environmental Protection Agency, Research Triangle Park, NC. EPA 452/R-11-003. April 2011.

In considering distributional statistics from epidemiological studies, we recognized that there are two types of study population-level metrics that are useful to consider in identifying the $PM_{2.5}$ concentrations most influential in generating the health effect estimates reported in the epidemiological studies. The most relevant information to consider is the distribution of health events (e.g., deaths, hospitalizations) occurring within a study population in relation to the distribution of $PM_{2.5}$ concentrations. However, in recognizing that access to event data can be restricted, and consistent with advice from CASAC on the second draft *Policy Assessment* we also considered the number of study participants within each study area as a reasonable surrogate for health event data.[2,3]

In the absence of health event and study population data, we also considered the distribution of $PM_{2.5}$ concentrations across study areas as representative of the $PM_{2.5}$ concentrations likely experienced by study participants, which are integral to the generation of effect estimates. This approach is particularly relevant for identifying the range of $PM_{2.5}$ concentrations most influential in generating the effect estimates in short-term exposure studies (e.g., time-series studies), but, as noted by CASAC, is more "complex" for long-term $PM_{2.5}$ exposure studies (Samet, 2010, p. 2). In short-term exposure studies, a similar number of events are likely occurring on a daily basis and, thus, information on the $PM_{2.5}$ air quality distributions provides a better approximation of a study participant's likely exposure in relation to the observed health effects as compared to long-term exposure studies where individual study participant's exposures change over the course of the study follow-up period (Samet, 2010, p. 2).

Specific analyses conducted, including tasks, assumptions, caveats, and processing methodologies are described in more details below. In summary, these analyses address the following:

- *Analysis 1* – Distributions of air quality and associated health event and population data from selected *long-term* epidemiological studies
- *Analysis 2* – Distributions of air quality and associated health event and population data from selected *short-term* epidemiological studies

General Data Processing

These analyses built on previous air quality analyses described in earlier EPA staff memoranda (Schmidt et al., 2010; Hassett-Sipple, et al., 2010).[4] For the current analyses, we

[2] Samet J. (2010, p. 2). Letter from Dr. Jonathan M. Samet, Chair, Clean Air Scientific Advisory Committee to the Honorable Lisa P. Jackson, Administrator, US EPA. CASAC Review of Policy Assessment for the Review of the PM NAAQS – Second External Review Draft (June 2010). September 10, 2010. Available at: http://yosemite.epa.gov/sab/sabproduct.nsf/264cb1227d55e02c85257402007446a4/CCF9F4C0500C500F8525779D0073C593/$File/EPA-CASAC-10-015-unsigned.pdf.

[3] A limitation of applying population data as a substitute for the number of the health events occurring in a study area is the recognition that baseline incidence rates of health events will vary across study areas.

[4] Schmidt M; Hassett-Sipple B; Rajan P (2010). $PM_{2.5}$ Air Quality Analyses. Memorandum to PM NAAQS review docket EPA-HQ-OAR-2007-0492-0131. July 22, 2010; Hassett-Sipple, B, Schmidt, M, and Rajan, P. (2010) Analyses of $PM_{2.5}$ Data for the PM NAAQS Review. Memorandum to the PM NAAQS review docket. Docket ID number EPA-HQ-OAR-2007-0492-0077. March 29, 2010. Available: http://www.epa.gov/ttn/naaqs/standards/pm/s_pm_2007_td.html.

requested population-level data from investigators of the six multi-city U.S. studies included in Analysis 2 of Schmidt et al. (2010). As described in Hassett-Sipple et al. (2010) these six studies were selected because they considered multiple locations representing varying regions across several seasons that provide evidence on the influence of climate and particle mixtures on health effects associated with long- or short-term $PM_{2.5}$ exposures. In addition, these multi-city studies considered relatively more recent air quality conditions (1999 to 2005).[5]

Specifically, the population-level data for these analyses originated from the authors of the following four multi-city long- and short-term $PM_{2.5}$ exposure epidemiological studies[6]:

Long-term $PM_{2.5}$ exposure studies:

- ACS Reanalysis II
  - Krewski D, Jerrett M, Burnett RT, Ma R, Hughes E, Shi Y, Turner MC, Pope AC III, Thurston G, Calle EE, Thun MJ. (2009). Extended Follow-Up and Spatial Analysis of the American Cancer Society Study Linking Particulate Air Pollution and Mortality. HEI Research Report 140, Health Effects Institute, Boston, MA. Available: http://pubs.healtheffects.org/view.php?id=315.

  - Burnett R (2010). Personal communication with Dr. Rick Burnett: Fw: Urgent Request for Information from your 2009 study. September 13, 2010. Document ID EPA-HQ-OAR-2007-0492-0304.

- Women's Health Initiative (WHI)
  - Miller KA, Siscovick DS, Sheppard L, Shepherd K, Sullivan JH, Anderson GL, Kaufman JD (2007). Long-term exposure to air pollution and incidence of cardiovascular events in women. N Engl J Med, 356: 447-458.

  - Miller K (2010). Personal communication with Kristin Miller: Re: Urgent Request for Information from your 2007 study. August 24, 2010. Document ID EPA-HQ-OAR-2007-0492-0306.

Short-term $PM_{2.5}$ exposure studies:

- Medicare Air Pollution Study (MCAPS)
  - Bell ML, Ebisu K, Peng RD, Walker J, Samet JM, Zeger SL, Dominic F (2008). Seasonal and regional short-term effects of fine particles on hospital admissions in 202 U.S. counties, 1999-2005. Am J Epidemiol, 168: 1301-1310.

[5] The network of $PM_{2.5}$ FRMs reporting to the AQS has been operational since 1999.

[6] We requested but did not receive population-level data from the investigators of the following 2 studies: Eftim SE; Samet JM; Janes H; McDermott A; Dominici F (2008). Fine Particulate Matter and Mortality: A Comparison of the Six Cities and American Cancer Society Cohorts With a Medicare Cohort. Epidemiology, 19: 209- 216., and Dominici F; Peng RD; Bell ML; Pham L; McDermott A; Zeger SL; Samet JL (2006). Fine particulate air pollution and hospital admission for cardiovascular and respiratory diseases. JAMA, 295: 1127-1134.

- Bell M (2010). Personal communication with Dr. Michelle Bell: Re: Urgent Request for Information from your 2008 study. August 30, 2010. Document ID EPA-HQ-OAR-2007-0492-0305.

- 112-Cities Mortality Study
  - Zanobetti A, Schwartz J (2009). The effect of fine and coarse particulate air pollution on mortality: A national analysis. Environ Health Perspect, 117: 1-40.

  - Zanobetti A (2010). Personal communication with Dr. Antonella Zanobetti; email to Pradeep Rajan, US EPA, OAQPS. September 15, 2010. Document ID EPA-HQ-OAR-2007-0492-0308, EPA-HQ-OAR-2007-0492-0308.1.

We requested that investigators from the three mortality studies noted above (Krewski et al., 2009; Miller et al., 2007; Zanobetti and Schwartz 2009) provide us with the total number of all-cause, non-accidental deaths and the total number of cohort members for each of the study areas (i.e., MSA, cities). In addition, we requested that investigators for one morbidity study noted above (Bell et al., 2008) provide us with the number of cardiovascular-related and respiratory-related hospitalizations and the total number of cohort members for each of the 202 study areas (i.e., counties).

To maintain continuity and consistency, the same air quality database included in prior analyses was utilized for the current analyses. The air quality data for this project originated from EPA's Air Quality System (AQS) database, the official repository of NAAQS-comparable ambient measurements. Data were extracted and processed as described in Hassett-Sipple et al. (2010).

In computing $PM_{2.5}$ concentrations corresponding with health event and population distributions for the four epidemiological studies noted above, the number of health events and study participants were combined for those study areas with identical $PM_{2.5}$ concentrations. A description of study-area specific $PM_{2.5}$ concentrations can be found in Attachment A of Hassett-Sipple et al. (2010). Additional data processing details are described in the analysis-specific descriptions below.

Analysis 1 – Distributions of air quality and associated health event and population data from selected long-term epidemiological studies

In the analysis documented here, we present distributions of health event (i.e., deaths, hospitalizations) and population (i.e. study participants) data from Krewski et al. (2009) and distributions of population data from Miller et al. (2007). Investigators provided us with MSA-specific counts of deaths and study participants for Krewski et al. (2009), and MSA-specific counts of study participants for Miller et al. (2007). These two long-term $PM_{2.5}$ exposure epidemiological studies provided a continuous measure of the density of the data from multi-city studies as a function of the annual mean concentrations [14.0 $\mu g/m^3$ (Krewski et al., 2009); 12.9

$\mu g/m^3$ (update of Miller et al. (2007) $PM_{2.5}$ data included in Curl (2009)[7]] across the cities in each study. Specifically, we considered the distributions of air quality data as well as distributions of event and study population data across air quality concentrations.

In the absence of event data we considered the distribution of population data as a surrogate for the occurrence of health events across study areas. Specifically, our analysis of the distributions of health events and population data from Krewski et al. (2009) examined the reliability of utilizing study population data as a surrogate for event data. Table 1 includes the years of air quality data, total study areas, and the annual $PM_{2.5}$ concentrations computed from the site-level 24-hour average $PM_{2.5}$ concentrations that correspond with study areas contributing to the 10^{th}, 25^{th}, and 50^{th} percentiles of the health event and study population data. Since the 10^{th}, 25^{th}, and 50^{th} percentiles for the health events and study population distributions are nearly identical for Krewski et al. (2009), we conclude that the distribution of population data can be a useful surrogate for event data, and in particular for Miller et al. (2007), for which we did not have health event data.

In Figure 1, the cumulative variable (i.e., frequency of deaths, study population) for Krewski et al. (2009) and Miller et al. (2009) is plotted on the y-axis as a percentage of the total study variable (i.e., as a percentage of the total study area-period count of health events or study population). The x-axis represents the annual mean concentration in units of $\mu g/m^3$.

Analysis 2 – Distributions of air quality and associated health event and population data from selected short-term epidemiological studies

This analysis built on previous air quality analyses described in Schmidt et al. (2010). In the analysis documented here, we present distributions of health events (i.e., deaths, hospitalizations) and population (i.e. study participants) data from two short-term $PM_{2.5}$ exposure studies (Zanobetti and Schwartz, 2009; Bell et al., 2008). The long-term mean $PM_{2.5}$ concentrations for Zanobetti and Schwartz (2009) and Bell et al. (2008) were 13.2 $\mu g/m^3$ and 12.9 $\mu g/m^3$, respectively.

Table 1. Summary Results for Select Long-and Short-term $PM_{2.5}$ Exposure Studies

Cite	Years of Air Quality Data	Study Areas	Population-Level Data	$PM_{2.5}$ ($\mu g/m^3$)[a] Distributional Statistics		
				10^{th}	25^{th}	50^{th}
Krewski et al. (2009)	1999-2000	116 US MSAs	Deaths	10.2	12.0	14.4
			Study Population	10.2	12.1	14.4

[7] Personal communication with Cynthia Curl, MESA Air Project Manager, University of Washington; email to Beth Hassett-Sipple, US EPA, OAQPS regarding request for PM air quality data. August 10, 2009. Document ID EPA-HQ-ORD-2007-0517-0113.

Miller et al. (2007)	2000	36 MSAs	Study Population	9.7	11.2	12.8
Zanobetti & Schwartz (2009)	1999-2005	99 US cities[b]	Deaths	10.3	12.5	14.5
			Study Population	10.3	12.5	14.5
Bell et al. (2008)	1999-2005	202 US counties	CV-related HA	9.8	11.5	13.5
			Respiratory-related HA	9.8	11.4	13.5
			Study Population	9.6	11.2	13.3

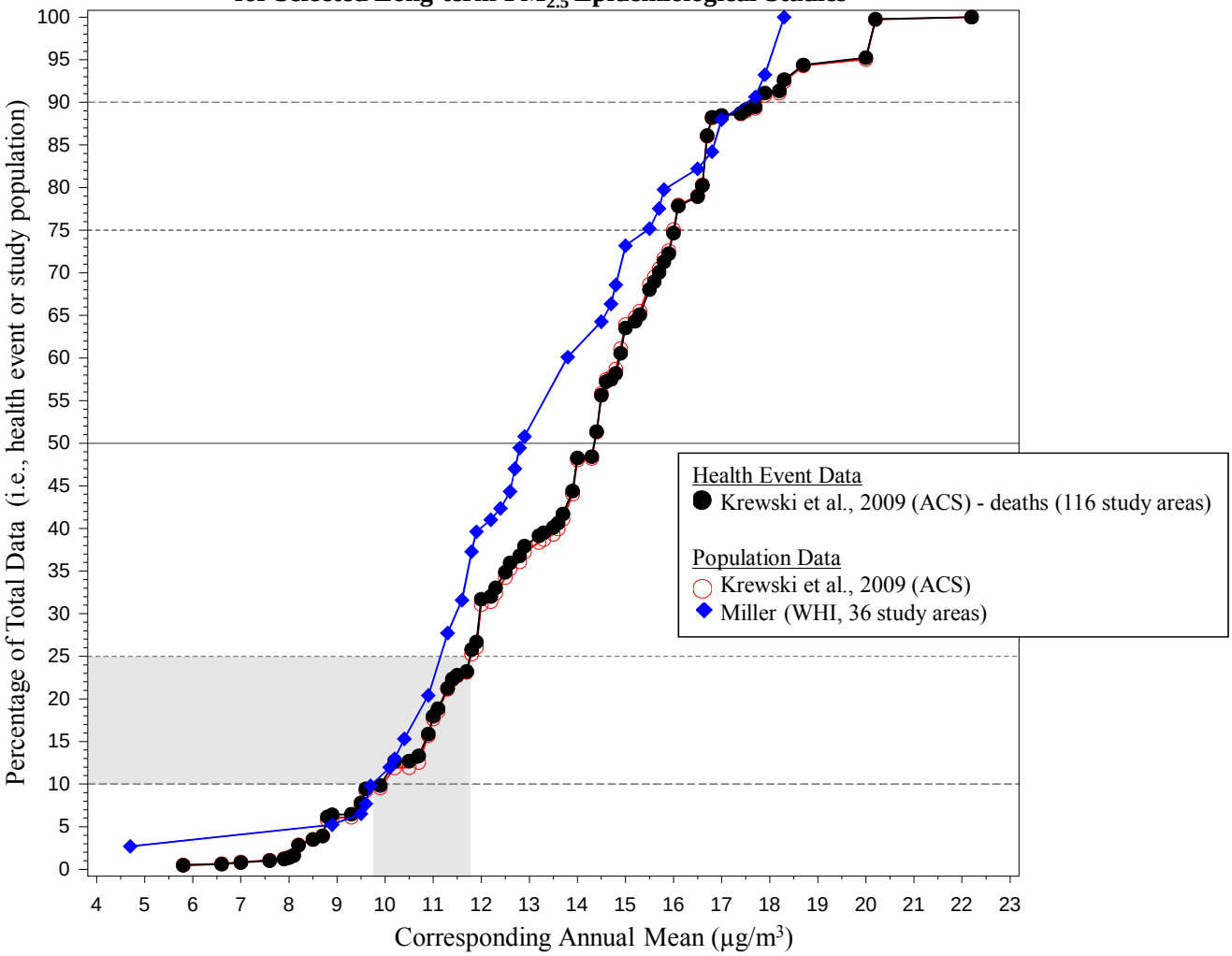
[a]Annual mean $PM_{2.5}$ concentration corresponding with study areas contributing to the 10^{th}, 25^{th}, and 50^{th} percentiles of the distribution of study population-level data.
[b]AQS data were available for 99 of 112 study areas included in Zanobetti and Schwartz (2009).

Investigators provided us with county- or city-specific counts of cardiovascular-related and respiratory-related hospitalizations (Bell et al., 2008) and deaths (Zanobetti and Schwartz, 2009). In addition, investigators for the Bell et al. (2008) study provided us with county-specific counts of study participants. For the Zanobetti and Schwartz (2009) study populations were estimated for each study by summing the county populations (per 2000 Census) for the cities where sites had valid contributing data for the area-period air quality metric estimates.

Table 1 includes for these two studies the years of air quality data, total study areas, and the annual $PM_{2.5}$ concentrations computed from the site-level 24-hour average $PM_{2.5}$ concentrations that correspond with study areas contributing to the 10^{th}, 25^{th}, and 50^{th} percentiles of the health event and population data. We note that as in the case of *Analysis 1*, the 10^{th}, 25^{th}, and 50^{th} percentiles for the health events and population distributions are nearly identical for Zanobetti and Schwartz (2009) and Bell et al. (2008), and thus provide further support for utilizing the distribution of population data as a surrogate for event data.

In Figure 2, the cumulative variable (i.e., frequency of deaths, cardiovascular and hospital admissions, study population) is plotted on the y-axis as a percentage of the total study variable (i.e., as a percentage of the total study area-period count of health events or study population). The x-axis represents the annual mean concentration in units of $\mu g/m^3$.

Figure 1. Distribution of Population-Level Data and Corresponding $PM_{2.5}$ Concentrations for Selected Long-term $PM_{2.5}$ Epidemiological Studies



In the absence of health event and study population data, the distribution of $PM_{2.5}$ concentrations across study areas could be used to represent the $PM_{2.5}$ concentrations likely experienced by study participants. However, we were unable to determine from the methodologies published for these studies how the air quality data were statistically weighted to account for variations in the availability of daily $PM_{2.5}$ measurements by study area. Consequently, we are unable to consider this type of information as part of our effort to identify the broader range of $PM_{2.5}$ concentrations that were most influential in generating the health effect estimates in epidemiological studies.

Figure 3 includes the distribution of health event data for Krewski et al. (2009), Zanobetti and Schwartz (2009), and Bell et al. (2008), respectively, and corresponding $PM_{2.5}$ concentrations. We also include the distribution of study participants for Miller et al. (2007) and corresponding $PM_{2.5}$ concentrations.

Figure 2. Distribution of Population-Level Data and Corresponding $PM_{2.5}$ Concentrations for Selected Short-term $PM_{2.5}$ Epidemiological Studies

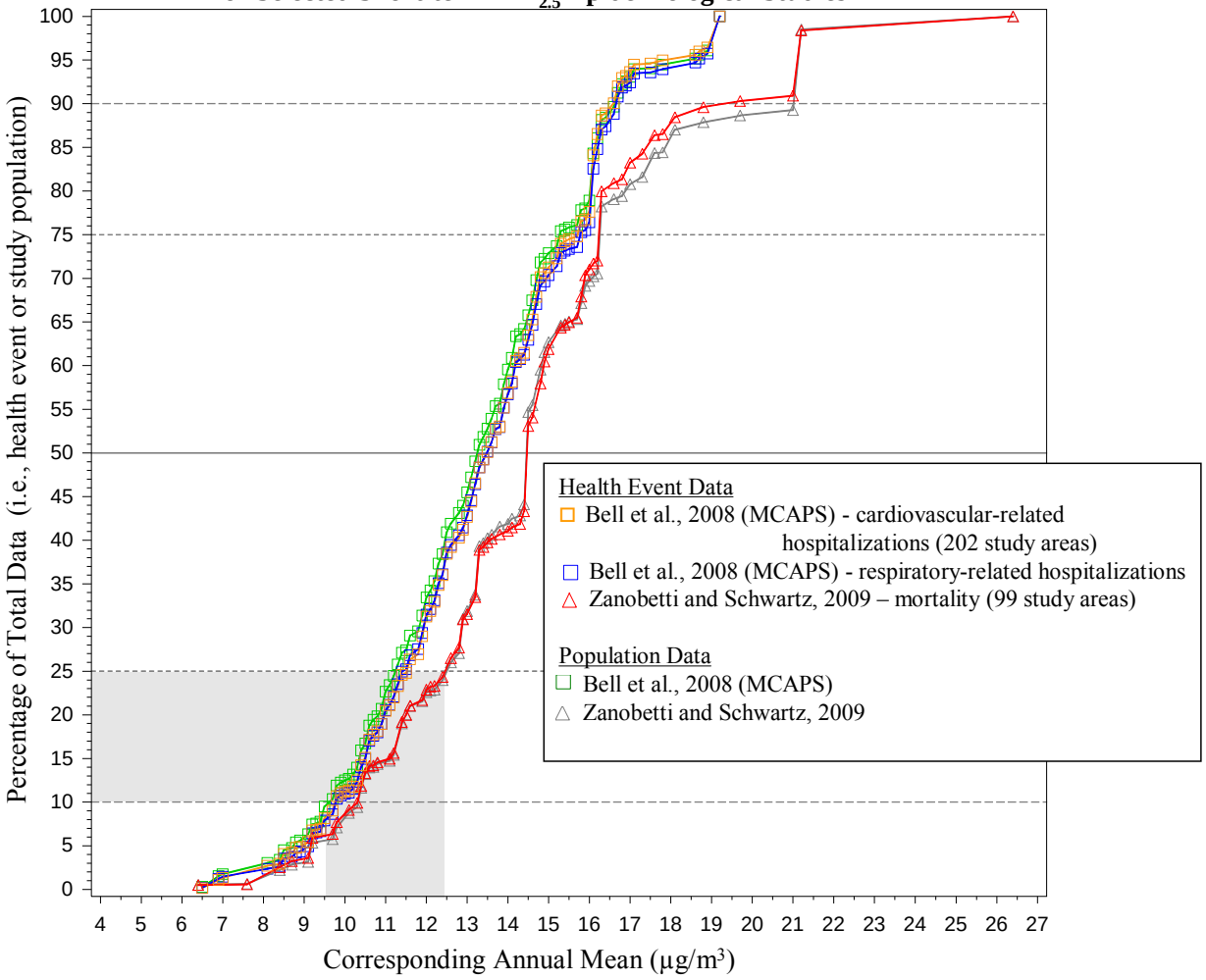


Figure 3. Distribution of Population-Level Data and Corresponding $PM_{2.5}$ Concentrations for Selected Multi-City Epidemiological Studies