Comments on EPA's Policy Assessment for PM2.5

Joel Schwartz Professor of Environmental Health Harvard TH Chan School of Public Health

p 2-49. Despite later referencing the Di 2019 PM model, the PA reports results for Di 2016 in most of its discussion about models, and comparing their predictions and not Di 2019, where CV R2 is 0.86 daily, 0.89 annually, and which was the source of data for the newer papers assessed in the ISA supplement providing new evidence on PM2.5 and health effects. Pleasse note that downscaler simplicity in fitting is not an advantage when the other models have already been fit and available. It certainly does not justify using a worse fitting model.

The statement that "Relatively weak performance in parts of the western U.S., possibly due to the sharp concentration gradients, complex terrain, low concentrations (and therefore signal-to-noise ratio), less dense monitoring, prevalence of wildfire, and challenges in satellite retrievals and CTM modeling (Di et al., 2016; Wang et al., 24 2018b; Hu et al., 2017; Kelly et al., 2019a)." again ignores the later model of Di 2019, used in many discussed studies. Di 2019 reports a cross validated (CV) R² of 0.80 daily and 0.85 annually for the Pacific region. That is not much weaker than national, and certainly an excellent fit. The PA's omission of more recent PM models overestimates uncertainty.

Di 2019 does not report weaker predictive ability at lower concentrations. It reports a) that the relationship between predicted and monitored values remains linear from 0 to $60 \ \mu g/m^3$ and that the uncertainty in the relationship was smallest between 4 and 15 $\ \mu g/m^3$. This is shown in the figure below from that paper, showing the spline fit and 95% CI. Clearly the uncertainty is low between 5 and 10.



Fig. 4. Relationship between monitored and predicted $PM_{2.5}$ at annual level. We regressed annual averaged $PM_{2.5}$ predictions from the ensemble model against annual averaged monitored $PM_{2.5}$ in a generalized additive model, with spline on the monitored $PM_{2.5}$. Dashed lines represent 95% confidence interval.

It can also be seen in the raw data, as shown in the below figure from Schwartz 2021 (Schwartz JD, Yitshak-Sade M, Zanobetti A, Di Q, Requia WJ, Dominick F, Mittleman MA. A self-controlled approach to survival analysis, with application to air pollution and mortality. Environment International, 2021, 157; 106861).



Fig. 1. Annual average Predicted Vs Measured PM2.5 (µg/m³) at Monitoring Locations in the U.S. The error around the line is similar at 5 and 12 µg/m³.

In comments on the previous PM2.5 proposed standard I submitted data where we fit our model predictions to $PM_{2.5}$ measurements at over 1900 monitoring stations. We compared the predicted annual average $PM_{2.5}$ at each location to the measurements. The difference is the error in the estimated concentration. To see how the error variance changes with $PM_{2.5}$ concentrations, we squared each error, and then smoothed the squared errors vs. the measured $PM_{2.5}$ concentrations at those monitors in those years. This estimates how the error variance changes with concentration. The results are on the plot below. They clearly show that the exposure error is **smaller** at concentrations below 12 µg/m³ and at a minimum at 6 µg/m³, not larger as the PA asserts. This makes sense, as there is more data in that range to estimate the predictions.



P 2-58 and following perform a bizarre analysis to justify the argument that the PM2.5 exposure in the epidemiology studies using multiple exposure models, including the Di et al exposure, produce lower effect estimates than the monitor level design value, and hence cannot be used to accurately judge the concentrations at which health effects are happening, or to set standards enforced at monitors. It again uses the Di 2016 exposure rather than the Di 2019 exposure paper. This is a strange choice since that model both had a better fit and was the one used in recent studies which has produced the low-exposure concentration-response curves that were cited by the ISA as showing effects below $10 \ \mu g/m^3$. But even stranger, rather than directly compare monitor measurements to predictions of these values at the monitoring locations, it computes the average PM2.5 concentration across all grid cells in all metropolitan areas in the US, and then compares that mean to the average monitor based maximum design values across all metropolitan areas in the U.S. This has nothing to do with whether the models are misrepresenting what is happening at the monitors, including the design monitor, or underpredicting exposure. The design value is the monitor value at the monitoring location that has the highest three-year average concentration in that AQMD. By definition, all the other monitoring locations in the same urban area had lower concentrations, and given that siting criteria tend to put monitors in more polluted parts of a metropolitan area, so would unmonitored locations. So why would an urban area wide average not be expected to have lower concentrations as well? This is simply not evidence of bias in the predictions of the hybrid models, i.e. that they tend to underestimate ambient concentrations. And averaging across the entire urban US will distort the association since not every urban area in the U.S. has monitors, and the unmonitored locations tend to have lower concentrations.

If you want to compare to the hybrid predictions to the design values, you should compare the hybrid predictions at the design value monitor to that monitor's readings. But comparing predictions at lower exposure locations to the design monitor values misrepresents the ability of those models to capture true ambient concentrations.

A fairer comparison is to compare predicted values and monitored values at all the monitors in the U.S. to see whether there is evidence of any bias, overall or at lower concentrations. As reported in Di 2019, we took our predictions at each EPA monitoring site in the U.S. and regressed them with the measured values at that monitor. The predictions were always from models that excluded the monitor being predicted. The results for daily values was an intercept of $0.625 \ \mu g/m^3$ and a slope of 0.956. This represents very little bias in the exposures. Of course, this relationship between monitored values and predictions at the monitors was for daily predictions. I have repeated the analysis for annual average monitor values and annual average predictions, which is most relevant for the annual standard, and is the exposure used in the cohort studies that used our exposure. The results indicate even less bias; the intercept was 0.077, and the slope was 0.969. There is clearly no noticeable bias here, and therefore no reason to believe that an effect seen at model predicted annual values of e.g. $8 \ \mu g/m^3$ is not an effect at that ambient concentration. This almost linear relationship with a slope indistinguishable from 1 and little if any bias is seen at all concentrations, as shown in the below figure, also from Di et al, 2019.



Monitored PM_{2.5} (μ g/m³)

The inaccurate and inappropriate assessment of bias in the way the policy assessment compares monitored and predicted values applies to the other hybrid models as well. It is not evidence of downward bias in predictions at all.

The Policy Assessment also states (p 2-61) that the monitors used for validation have few monitors few monitors in less densely populated areas where concentrations are low. In the case of Di et al 2019, the monitors used included the National Park Service IMPROVE network, which is generally in low populated areas, the CASTNET network, also mostly in rural areas, the SEARCH network, which has suburban monitors, and data from the Southern California Children's Health Study, also including data from suburban areas. As the confidence intervals above show, one has to get to concentrations below 5 μ g/m³ before the uncertainty starts to widen.

P2-61 also states ", performance evaluations for the methods are weighted toward densely monitored urban areas at the scales of representation of the monitoring networks. Predictions at different scales or in sparsely monitored areas are relatively untested. Second, studies have reported heterogeneity in performance with relatively weak performance in parts of the western U.S., at low concentrations, at greater distance to monitors, and under conditions where the reliability and availability of key input datasets (e.g., satellite retrievals and air quality modeling) are limited." Again, this is not true for the Di 2019 model, and why the PA ignores this is mystifying, particularly as it is the source of exposure for many recent studies, including studies focused on low concentrations. The studies below all used this exposure.

Ward-Caviness CK, Weaver AM, Buranosky M, Pfaff ER, Neas LM, Devlin RB, Schwartz J, Di Q, Cascio WE, Diaz-Sanchez D. Associations between long-term fine particulate matter exposure and mortality in heart failure patients. J American Heart Association, 2020, Pages: e012517. (NOTE THAT THIS IS AN EPA STUDY);

Qiu X, Wei Y, Wang Y, Di Q, Sofer T, Abu Awad Y Schwartz J. Inverse probability weighted distributed lag effects of short-term exposure to PM2.5 and ozone on CVD hospitalizations in New England Medicare participants-exploring the causal effects. Environmental Research, 2020:182:109095.

Schwartz J, Wei Y, Yitshak-Sade M, Di Q, Dominici F, Zanobetti A. A National Difference in Differences Analysis of the Effect of PM2.5 on Annual Death Rates. Environmental Research, 2020 194:110649, DOI: 10.1016/j.envres.2020.110649 PMID: 33385394

Wei Y, Wang Y, Wu X, Di Q, Shi L, Koutrakis P, Zanobetti A, Dominici F and Schwartz J. Causal effects of air pollution in Massachusetts. Am J Epidemiol. 2020; 11:1316-1323 PMID: 32558888.

Yitshak-Sade M, Nethery R, Schwartz J, Mealli F, Dominici F, Di Q, Abu Awad Y, Ifergane G, and Zanobetti A. PM2.5 and Hospital Admissions Among Medicare Enrollees with Chronic Debilitating brain disorders. Aug 2020. The Science of the Total Environment. 2021:756:142524.

Shi L, Wu X, Danesh Yazdi M, Braun D, Abu Awad Y, Wei Y, Liu P, Wang Y, Schwartz J, Dominici F, Kioumourtzoglou M-A, Zanobetti A. Long-term effects of PM2·5 on neurological disorders in the American Medicare population: a longitudinal cohort study. Lancet Planet Health 2020, https://doi.org/10.1016/S2542-5196(20)30227-8.

Qiu X, Fong KC, Shi L, Papatheodorou S, Di Q, Just A, Kosheleva A, Messerlian C, Schwartz J. Prenatal exposure to particulate air pollution and gestational age at delivery in Massachusetts neonates 2001–2015. Environmental Epidemiology, 2020, DOI: 10.1097/EE9.00000000000113.

Cserbik D, Chen J-C, McConnell R, Berhane K, Sowell ER, Schwartz J, Hackman DA, Kan E, Fan CC, Herting MM. Fine particulate matter exposure during childhood relates to hemispheric-specific differences in brain structure. Environment International, 2020,143, https://doi.org/10.1016/j.envint.2020.105933

Wu X, Braun D, Schwartz J, Kioumourtzoglou M-A, Dominici F. Evaluating the impact of long-term exposure to fine particulate matter on mortality among the elderly. Scientific Advances, 10.1126/sciadv.aba5692 (2020).

Schwartz J, Di Q, Requia W, Dominici F, Zanobetti A. A Direct Estimate of the impact of PM2.5, NO2, and O3 Exposure on Life Expectancy using Propensity Scores. Epidemiology, 2021 Jul 1;32(4):469-476. doi: 10.1097/EDE.00000000001354

DeSouza P, Braun D, Parks RM, Schwartz J, Dominici F, Kioumourtzoglou M-A. Nationwide study of Short-term exposure to fine particulate matter and cardiovascular hospitalzations among Medicaid enrollees. Epidemiology, 2021:32, 6-13.

Zhang S, Breitner S, Cascio W, Devlin RB, Neas LM, Ward-Caviness C, Diaz-Sanchez D, Kraus WE, Hauser ER, Schwartz JD, Petters A, Schneider A. Association between short-term exposure to ambient fine particulate matter and myocardial injury in the CATHGEN cohort. Environmental Pollution, 2021, DOI: 10.1016/j.envpol.2021.116663. (ALSO AN EPA STUDY)

Danesh Yazdi M, Wang Y, Di Q, Wei Y, Requia WJ, Shi L, Sabath MB, Dominici F, Coull BA, Evans JS, Koutrakis P, Schwartz JD. Long-Term Association of Air Pollution and Hospital Admissions Among Medicare Participants Using a Doubly Robust Additive Model. Circulation, 2021. DOI: 10.1161/CIRCULATIONAHA.120.050252.

Alexeeff SE, Deosaransingh K, Liao NS, Van Den Eeden SK, Schwartz J, Sidney S. Particulate Matter and Cardiovascular Risk in Adults with Chronic Obstructive Pulmonary Disease. Am J Respir Crit Care Med. 2021 Mar 4. doi: 10.1164/rccm.202007-2901OC. Epub ahead of print. PMID: 33662228

Son J-Y, Sabath MB, Lane KJ, Miranda ML, Dominici F, Di Q, Schwartz J, Bell ML. Long-term exposure to PM2.5 and mortality for the older population: Effect modification by residential greenness. 2021, Epidemiology, DOI: <u>10.1097/EDE.00000000001348</u>

Ward-Caviniess CK, Danesh Yazdi M, Moyer J, Weaver AM, Cascio WE, Schwartz JD, Diaz-Sanchez D. Long-Term Exposure to Particulate Air Pollution Is Associated With 30-Day Readmissions and Hospital Visits Among Patients With Heart Failure. J American Heart Association, 2021.

Wei, Y., M. Yazdi, Q. Di, W. Requia, F. Dominici, A. Zanobetti and J. Schwartz (2021). Emulating causal dose-response relations between air pollutants and mortality in the Medicare population. Environmental Health 20 Article number: 53 (2021).

Schwartz JD, Yitshak-Sade M, Zanobetti A, Di Q, Requia WJ, Dominick F, Mittleman MA. A selfcontrolled approach to survival analysis, with application to air pollution and mortality. Environment International, 2021, 157; 106861.

Danesh Yazdi M, Wang Y, Di Q, Requia WJ, Wei Y, Shi L, Sabath MB, Dominici F, Coull. BA, Evans JS, Koutrakis P, Schwartz. JD. Long-term Effect of Exposure to Lower Concentrations of Air Pollution on Mortality among Medicare Participants and Vulnerable Subgroups. Lancet Planetary Health, 2021, **5**(10): e689-e697.

Lu W, Hackman DA, Schwartz J. Ambient air pollution associated with lower academic achievement among US children A nationwide panel study of school districts. Environmental Epidemiology. 2021, 00:e174

Shi L, Steenland K, Li H, i Liu P, Zhang Y, Lyles R, Requia W, Ilango S, Chang H, Wingo T, WeberR, Schwartz J. A national cohort study (2000-2018) of long-term air pollution exposure and incident dementia in older adults in the United States. Nature Communications, 2021.

Note that many of these studies focused on low exposure levels, so these are very relevant to reviewing the standard, and are indeed discussed in the supplemental ISA as providing health evidence below current standards, and even as low as $5 \,\mu g/m^3$.

As to how well the Di et al 2019 model did, the value reported as spatial R2, which is the R2 for annual averages, was 0.85 in the Pacific region, Vs 0.89 overall. While that is lower, it is not relatively weak. The region with the lowest R2 for annual average predictions was the mountain states, which had an R2 of 0.77 which is still quite good. EPA routinely does risk assessments using CAMx or CMAQ models whose R2 hover around 0.5, for example. The already reported figure below also puts the lie to the assertion that the models have more error at lower concentrations.



Monitored PM2.5

P 3-6 and elsewhere the PA states that lowering the annual average standard will also reduce the number of high days. While this is true, it does not mean that the number of days when PM2.5 concentrations are higher than levels which cause significant harm, including death, will be minimized. This is very problematic. To illustrate this, I have taken all the EPA monitors, and restricted to years where the annual average was below 8 μ g/m³. I then computed the average number of days when PM2.5 exceeded 20 μ g/m³ at each monitor. On average, each monitor had days exceeding 20 μ g/m³ 1.29% of the time, which corresponds to 4.7 days per year. To see the importance of this, consider the below figure from Wei 2020. It clearly demonstrates that daily mortality is associated with daily PM2.5 concentrations at levels even below 20 μ g/m³, let alone higher. Hence an annual standard of 8 μ g/m³ would not protect people from days above 20 μ g/m³, where the risk of mortality is not low. To make that clear, compare a probability of dying of 10⁻⁶ per **day** with EPA's National Contingency Plan (40 C.F.R. § 300.430(d)(1)) which states that

the range of acceptable lifetime risks (of developing cancer) for carcinogens should be set between 1 in 10,000 and 1 in a million over a **70-year lifetime.**



Or consider that in 2017, we published a paper examining over 22 million deaths in over 39,000 ZIP Codes of participants in Medicare. Of those deaths, 95% occurred when daily PM_{2.5} was below 25 μ g/m³, well below the daily NAAQS of 35 μ g/m³.(Di , Wang et al. 2017). Another recent publication looked at daily PM_{2.5} and daily deaths in a study of 652 cities world-wide. It also examined how the association varied with PM_{2.5} concentrations. They reported that exposure to PM_{2.5} concentrations lower than 20 μ g/m³ was significantly associated with daily deaths (Liu, Chen et al. 2019).

In sum, an annual standard of 8 μ g/m³ would allow days where concentrations in the U.S. exceeded 20 μ g/m³ 1.29% of the time, which would result in hundreds to thousands of deaths that could be avoided by lowering the daily standard as well.

P 3-23. Some studies missed by the draft ISA supplement are important. For example, Danesh Yazdi 2021 updated the Di et al. Medicare cohort to include follow-up until 2016 and restricted to people exposed to PM2.5 < 12 μ g/m³. She reported a significant association with mortality in a doubly robust causal model. Schwartz 2021 used a novel self-controlled design which compared exposure in the year subjects died to exposure of the same subject in all eligible years and reported a stronger association when the analysis was restricted to concentrations below 12 μ g/m³. Notably, this analysis was stratified by individual, with no contrast between people. Hence all individual level fixed or slowly varying covariates, measured or unmeasured, cannot confound in this study design.

Another study of Wei and coworkers (2021) examined mortality risks in 74 million Medicare participants. Controlling for co-pollutant exposures and using a propensity score approach they reported significant associations with PM2.5 down to the lowest exposure level. The effects are shown below.



Specifically, the RRs of mortality associated with chronic exposure to PM2.5 was 1.022 [95% confidence interval (CI), 1.018–1.025] at 6.60 μ g/m³ (the 2nd decile group) compared to the lowest decile of exposure. The PA should also note that effects below 8 μ g/m³ are supported by the studies of Pinault, Wei (2020), Wei (2021), Wang (2020), Christidis (2919), Pappen (2019), and Hayes (2020).

Also, the statement that "a few studies use statistical techniques to reduce uncertainties related to potential confounding" seems to underestimate the number of such studies. In particular, the difference in differences studies use a dummy variable for every Zip code, census tract, or city (depending on the study) that removes all unmeasured confounding that varies across geographic area, whether measured or not. The three studies of Wang, Schwartz, and Yitshak-Sade, cited by the supplemental ISA, used a dummy variable for every ZIP code (census tract in the paper of Wang) to control for all neighborhood SES, green space, availability of fresh food, and other characteristics, measured or unmeasured. An additional paper by me, which is not included in the supplemental ISA, fit a difference in differences analysis nationwide in the Medicare data, and reached the same conclusion (Schwartz J, Wei Y, Yitshak-Sade M, Di Q, Dominici F, Zanobetti A. A National Difference in Differences Analysis of the Effect of PM2.5 on Annual Death Rates. Environmental Research, 2020 194:110649, DOI: 10.1016/j.envres.2020.110649 PMID: 33385394.). That paper is attached to these comments for consideration. Also, the difference in differences study of Kioumourtzoglou (2015) looked at annual PM2.5 and mortality separately in each of 81 cities in the Medicare population. They fit a separate time trend in each of the cities they examined to capture city-specific time trends in mortality. Hence that controls for time trends in omitted covariates that may differ by city, as well as variables that may differ across cities.

Two additional studies provide assurances about potential confounding by unmeasured covariates. Abu Awad et al. performed an analysis restricted to people who were **never**

exposed above 12 μg/m³. They analyzed Medicare beneficiaries who had moved, and controlled for Zip code of origin. By only comparing people who had moved from the same location, they again controlled for all neighborhood level covariates, measured or unmeasured. They found strong associations between PM_{2.5} exposures in that range and annual death rates among the Medicare beneficiaries(Abu Awad, Di et al. 2019). Moreover, since the association was with **change** in exposure after moving, and stratified on ZIP code of initial residence, the association cannot reflect past exposures, which were the same for all the movers from the same ZIP code. Another study that controlled for unmeasured covariates was my recent paper, where individuals in the Medicare cohort who died were matched to themselves in different years when they did not die clearly avoids such confounding by design (*Schwartz JD, Yitshak-Sade M, Zanobetti A, Di Q, Requia WJ, Dominick F, Mittleman MA. A self-controlled approach to survival analysis, with application to air pollution and mortality. Environment International 157 (2021) 106861.*). This study, by definition, controlled by matching for all fixed or slowly varying individual characteristics. In summary, the number of studies that used novel methods to reduce the chance of unmeasured confounding is more than "a few".

P- 3-24. The PA states that "There was also some limited evidence indicating that the slope of the concentration-response (C-R) function may be steeper (supralinear) at lower concentrations for cardiovascular mortality (U.S EPA, 2021a, section 2.2.3.2)."

The evidence for steeper C-R at lower concentrations is not limited as shown in the study of Vodonos et al. (*Vodonos A, Abu Awad Y, Schwartz J. The concentration-response between long-term PM2.5 exposure and mortality; A meta-regression approach. Environmental Research, 2018, 166:677-689. doi: 10.1016/j.envres.2018.06.021. PMID: 30077140*) This was a meta-analysis that examined over 100 effect estimates from 52 cohorts, 14 of which had mean exposures below 10 μ g/m³, and overall the evidence supported a steeper slope at lower concentrations, as seen in the figure below, which shows the estimated slope Vs the mean concentration from that meta-regression.

Not included in that meta-analysis are the two studies of Wei. The first, show above shows a significant difference between a PM2.5 of 6.6 μ g/m3 and the lower concentrations in the first decile of exposure. The other study used sequential truncation of the data to assess the effect size at concentrations below certain thresholds. That result is shown below. Clearly the slope is higher at lower concentrations.



P 3-49. The PA states that there are fewer new studies of acute effects of PM2.5 and mortality but neglects to mention that one of them examined daily $PM_{2.5}$ and daily deaths in 652 cities world-wide. Surely that is better than 3 studies of 10 cities each. Moreover, It reported that exposure to $PM_{2.5}$ concentrations lower than 20 µg/m³ was significantly associated with daily deaths (Liu, Chen et al. 2019).

P 3-49. The PA treats PM2.5 related deaths from Cardiovascular causes and deaths from all causes as if they were independent outcomes to be assessed separately. They are not. Simple logic indicates that they must be connected. If PM2.5 causes increased deaths from cardiovascular deaths, then it **must** cause increased deaths from all causes, which contains cardiovascular. Hence the cause specific studies should be considered in drawing causal conclusions about total mortality. Similarly, if PM2.5 is causal for total deaths, then those people **must** have died of something. There must be specific causes of death that are caused by PM2.5. Since the major cause-specific deaths that have been associated with PM2.5 are cardiovascular deaths, respiratory deaths, and lung cancer deaths, a causal association with all-cause mortality means that PM2.5 **must** be causal for at least one of those causes of death. Moreover, a simple calculation of attributable risk from the evidence for all-cause mortality (that is, simply (RR-1)/RR for the mean concentration of PM2.5) produces a number of attributable deaths that cannot be due to any reasonable relative risk for diseases that are less common causes of death. It could only be due to increased risk from diseases with high mortality burdens, such as cardiovascular disease or cancer. Therefore, the discussion of

causality for cause-specific deaths should acknowledge this, and cite as evidence the causality with all-cause deaths, and vice versa.

P 3-64 misrepresents what epidemiology studies provide. It is true that they examine associations across a range of exposures, it is not true that they "do not identify the specific exposures that can lead to the reported effects". Multiple studies have used splines to estimate dose response curves, which can identify thresholds, and identify effects at given exposures. They also provide confidence intervals for each exposure, so one can identify the uncertainty about the effect size at e.g. 7 μ g/m³. In most of those studies, the confidence interval did not include zero. Some studies (e.g. Pinault) have fit piecewise linear models with multiple inflection points to identify if there is a threshold, i.e. a point below which the effect is zero. They have not found such a point. Other studies, such as the Wei et al paper discussed above, have excluded all observations above a series of values, and assessed whether there was a significant association between PM2.5 and death when restricted to observations below that value. Even when restricted to observations with annual averages below 6 µg/m³, they found a significant association with mortality. The Wei causal dose response paper divided PM2.5 into decile ranges and found a significant increase in risk of death at mean concentrations of 6.6 $\mu g/m^3$ compared to concentrations in the lowest decile. None of these approaches found a threshold. The epidemiology studies do not fail to identify discernable threshold because "they do not identify the specific exposures that can lead to the reported effects"; they fail to identify a discernable threshold because it does not exist.

P 3-65 seeks to focus on the mean concentration in studies stating. "there is significantly greater confidence in the magnitude and significance of observed associations for the part of the air quality distribution corresponding to where the bulk of the health events in each study have been observed, generally at or around the mean concentration. This is the case both for studies of daily PM2.5 exposures and for studies of annual average PM2.5 exposures." This distorts the meaning of the epidemiology studies and represents a lack of understanding of both epidemiology and statistics. First, there is no reason to rely on ad hoc poorly justified arguments about focusing on the mean to reduce uncertainty when the statistical tools used in the studies provide mathematically proven methods to compute confidence intervals, which take into account the number of observations. The elevated risk of death at an annual concentration of 6.6 μ g/m³ in the Wei et al. paper comes with an appropriately computed confidence interval, which is far from including the null. This despite it being only the 20th percentile of the exposure distribution. The spline models and the models that excluded higher observations also have confidence intervals. While it is true they are smallest near the mean, this does not mean that they were not small elsewhere. These models correctly estimate the uncertainty, including taking into account how many events occurred at that exposure. Moreover, while the mean is a good measure of central tendency for Gaussian data, such as annual PM2.5, it is not for skewed data such as daily PM2.5, where most of the data is below the mean, and a Median is a better measure. And remember that half the observations for Gaussian data occur below the mean, so studies are hardly uninformative in that range.

Second, the comment fails to understand that some of these studies had very large sample sizes, so that there were sufficient observations even far from the mean exposure. The Medicare cohort, for example, had 637 million person-years of observation. Hence there were over 63 million observations at the tenth percentile or lower. This is a very large number for an epidemiology study and clearly would have enough health effects to generate confidence if it were a separate study on its own. If the PA wants to argue about whether there are enough observations to be confident in, it should discuss the absolute number of observations in different exposure ranges in studies, not the proportion.

Third, the distribution of exposures even in smaller cohorts is wide enough that there are many observations and deaths occurring at concentrations well below the mean. Also, the choice of CBSA averages to compare to the maximum monitor seems chosen to maximize the difference by include large areas in the mean. CBSA's are often quite large. Finally, the reason to rely on confidence intervals and not just arguments about "where most of the observations are" is that confidence in an association depends not just on the number of observations. It depends on the strength of the association and other factors as well. This is accounted for in the confidence intervals, but not in the PA's approach. This argument to focus on the mean value in studies has no merit.

P 3-66. Again, large studies can provide strong support well below the mean. The support depends on the number of observations in that range, not the percentage.

P 3-67. The observation that the maximum is higher than the mean is truly revolutionary.

P 3-68. While it is true that the maximum monitor value is higher than the population average, it is also true that some people live near the maximum monitor. Hence, if a study reports increased death rates in a neighborhood where the concentrations are e.g. 8 μ g/m³, then failure to set a standard at 8 μ g/m³ will result to increased risk to people living near the maximum monitor, and at every other location where the exposure is above 8 μ g/m³, even if a maximum monitor value of e.g. 10 μ g/m³ meant that most people in that urban area were exposed to concentrations below 8 μ g/m³. Hence relying on most people being exposed below the maximum monitor value does not protect the heath of many people who are not. Again, this is based on the false belief that an epidemiology study only evaluates health effects at the mean exposure. The ability of hybrid exposure models to capture effects in suburban areas where exposure is lower is directly relevant to determining whether people in densely populated areas with currently higher exposures (e.g. the maximum monitor value) will benefit from a reduction in those exposures.

P 3-72. Table. When describing the Pinault study, this should note that Eighty percent of the population was exposed to concentrations below 8.8 μ g/m³.

P 3-79. The Puett studies rely on exposure estimates from the Yanoski model, and the model performance evaluation is in his studies not the Puett study. That is no reason for excluding it.

P 3-80. That studies using hybrid exposure models can capture whether there are health effects at low concentrations by providing evidence in suburban and rural areas is a key advantage of those studies. To argue that their mean exposure (irrelevant to the C-R curves) is therefore lower than the exposure at the highest monitor and therefore allowing higher exposure at the highest monitor is protective, turns logic on its head. The ability of hybrid exposure models to capture effects in suburban areas where exposure is lower is directly relevant to determining whether people in densely populated urban areas with currently higher exposures (e.g. the maximum monitor value) will benefit from a reduction in those exposures. Consider that in the Wei study, we found there is an elevated risk of death at 6.6 μ g/m³ in a study whose mean was 9.85 μ g/m³. Clearly this illustrates the poverty of interpreting the study as showing effects at it mean concentration. It also means that reducing exposure at the highest monitor from 10 μ g/m³ to 6.6 μ g/m³ would improve the health of people in that urban area who were exposed to 10, or 9, or 8, or 7 μ g/m³ by forcing all those exposures downward. To describe that study as estimating effect at 9.85 μ g/m³, and argue that a maximum monitor value of e.g. 11 or 12 μ g/m³ would suffice to protect people against the exposures seen in that study is ridiculous.

P3-86. Table 3-5 is missing the study of Schwartz, Fong, et al, Schwartz J, Fong K, Zanobetti A. A National Multicity Analysis of the Causal Effect of Local Pollution, NO2, and PM2:5 on Mortality. Environmental Health Perspectives. 2018, 126(8): doi.org/10.1289/EHP2732 PMID: 30235421. This is the largest US time series study, covering 135 cities, which also used 3 different types of causal modeling, and controlled for NO2 exposure in evaluating the effects of PM2.5. They report that on days below 25 μ g/m³ they found a significant increase in daily deaths associated with PM2.5, with an effect size per 10 μ g/m³ of 0.83% in daily deaths (95% Cl 0.39%, 1.27%). It is also missing Liu, C., R. Chen, F. Sera, A. M. Vicedo-Cabrera, Y. Guo, S. Tong, M. Coelho, P. Saldiva, E. Lavigne, P. Matus, N. Valdes Ortega, S. Osorio Garcia, M. Pascal, M. Stafoggia, M. Scortichini, M. Hashizume, Y. Honda, M. Hurtado-Díaz, J. Cruz, B. Nunes, J. Teixeira, H. Kim, A. Tobias, C. Íñiguez, B. Forsberg, C. Åström, M. Ragettli, Y.-L. Guo, B.-Y. Chen, M. L. Bell, C. Y. Wright, N. Scovronick, R. M. Garland, A. Milojevic, J. Kyselý, A. Urban, H. Orru, E. Indermitte, J. J. K. Jaakkola, N. R. I. Ryti, K. Katsouyanni, A. Analitis, Z. Z., J. Schwartz, J. Chen, T. Wu, A. Cohen, A. Gasparrini and H. Han (2019). "Ambient Particulate Air Pollution and Daily Mortality in 652 Cities." New Engl J Med 381(8): 705-715. which reports significant associations when restricted to observations below 20 μ g/m³.

Tables 3-6 and 3-7 is missing a number of important studies of long-term exposure, including: *Abu Awad, Y., Q. Di*, *Y. Wang, C. Choirat, B. Coull, A. Zanobetti and J. Schwartz (2019). "Change in PM2.5 exposure and mortality among Medicare recipients."* <u>Environmental Epidemiology</u> **3**: e054

Ward-Caviness CK, Weaver AM, Buranosky M, Pfaff ER, Neas LM, Devlin RB, Schwartz J, Di Q, Cascio WE, Diaz-Sanchez D. Associations between long-term fine particulate matter exposure and mortality in heart failure patients. J American Heart Association, 2020, Pages: e012517.

Schwartz J, Wei Y, Yitshak-Sade M, Di Q, Dominici F, Zanobetti A. A National Difference in Differences Analysis of the Effect of PM2.5 on Annual Death Rates. Environmental Research, 2020 194:110649,DOI: 10.1016/j.envres.2020.110649 PMID: 33385394.

Schwartz J, Di Q, Requia W, Dominici F, Zanobetti A. A Direct Estimate of the impact of PM2.5, NO2, and O3 Exposure on Life Expectancy using Propensity Scores. Epidemiology, 2021 Jul 1;32(4):469-476. doi: 10.1097/EDE.00000000001354

Alexeeff SE, Deosaransingh K, Liao NS, Van Den Eeden SK, Schwartz J, Sidney S. Particulate Matter and Cardiovascular Risk in Adults with Chronic Obstructive Pulmonary Disease. Am J Respir Crit Care Med. 2021 Mar 4. doi: 10.1164/rccm.202007-2901OC. Epub ahead of print. PMID: 33662228

Son J-Y, Sabath MB, Lane KJ, Miranda ML, Dominici F, Di Q, Schwartz J, Bell ML. Long-term exposure to PM2.5 and mortality for the older population: Effect modification by residential greenness. 2021, Epidemiology, DOI: <u>10.1097/EDE.00000000001348</u>

Wei, Y., M. Yazdi, Q. Di, W. Requia, F. Dominici, A. Zanobetti and J. Schwartz (2021). "Emulating causal dose-response relations between air pollutants and mortality in the Medicare population." Environmental Health 20 Article number: 53 (2021).

Schwartz JD, Yitshak-Sade M, Zanobetti A, Di Q, Requia WJ, Dominick F, Mittleman MA. A selfcontrolled approach to survival analysis, with application to air pollution and mortality. Environment International, 2021, 157; 106861.

Danesh Yazdi M, Wang Y, Di Q, Requia WJ, Wei Y, Shi L, Sabath MB, Dominici F, Coull. BA, Evans JS, Koutrakis P, Schwartz. JD. Long-term Effect of Exposure to Lower Concentrations of Air Pollution on Mortality among Medicare Participants and Vulnerable Subgroups. Lancet Planetary Health, 2021, **5**(10): e689-e697

This is a lot of evidence to omit, much of it showing effects at low concentrations, and much of it causal modeling.

P 3-98. The PA reports the empirical result that the average of all monitors in the Atlanta area is smaller than the value at the highest monitor. An impressive finding. This is again used to call into question the use of the mean value of exposure in studies using the hybrid models (which include unmonitored areas where pollution is lower) as demonstrating levels at which effects are seen. The use of averaged values in a study to estimate levels at which effects are seen is not the way to determine levels at which effects are seen. Rather, the concentration-response curves reported, the sequential truncation models reported, and the decile comparisons reported in Wei are the appropriate way. Nor, as noted above, can this approach conclude that higher concentrations at the maximum monitor are protective since most areas are lower, since the maximum monitor tends to be in densely populated areas with lots of people exposed to those levels. Finally, Figure 3-7, illustrating the size of the Atlanta CBSA compared to the state of Georgia again illustrates the problem of comparing the maximum monitor to values in a large area that is a substantial fraction of the entire state. Indeed, while the highest monitor average

was 10.4 μ g/m³, the other monitors in Atlanta report levels of 9.9 μ g/m³ and 9.3 μ g/m³, which is not that different from the design value monitor.

P 3-101. Again, the discussion of the design value being 40-50% higher than the large area or study average seems to be here to justify the argument that a higher design value adequately protects because it results in mean exposures that are much lower, ignoring that there are many people living in areas at or near the design value. It also seems designed to imply that the mean value in a study is the value at which it is detecting health effects, which is false.

P 3-102 again asks the wrong question. The question is not "What are the mean PM2.5 concentrations reported by key epidemiology studies". The appropriate question is what are the concentrations of PM2.5 at which key studies provide evidence of an effect? As noted above, this is quite different, since, e.g. the Wei study provides highly significant results at 10 μ g/m3 in a study whose mean was 9.8 μ g/m³. Similar findings are presented by many other of the cited studies. Further, the tables seem to omit key studies with lower concentrations. These include, for short-term studies:

Schwartz, Fong, et al, **Schwartz J**, Fong K, Zanobetti A. A National Multicity Analysis of the Causal Effect of Local Pollution, NO2, and PM2:5 on Mortality. Environmental Health Perspectives. 2018, 126(8): doi.org/10.1289/EHP2732 PMID: 30235421 and

Liu, C., R. Chen, F. Sera, A. M. Vicedo-Cabrera, Y. Guo, S. Tong, M. Coelho, P. Saldiva, E. Lavigne, P. Matus, N. Valdes Ortega, S. Osorio Garcia, M. Pascal, M. Stafoggia, M. Scortichini, M. Hashizume, Y. Honda, M. Hurtado-Díaz, J. Cruz, B. Nunes, J. Teixeira, H. Kim, A. Tobias, C. Íñiguez, B. Forsberg, C. Åström, M. Ragettli, Y.-L. Guo, B.-Y. Chen, M. L. Bell, C. Y. Wright, N. Scovronick, R. M. Garland, A. Milojevic, J. Kyselý, A. Urban, H. Orru, E. Indermitte, J. J. K. Jaakkola, N. R. I. Ryti, K. Katsouyanni, A. Analitis, Z. Z., J. Schwartz, J. Chen, T. Wu, A. Cohen, A. Gasparrini and H. Han (2019). "Ambient Particulate Air Pollution and Daily Mortality in 652 Cities." <u>New Engl J Med</u> **381**(8): 705-715.

For long-term studies these include the studies I mentioned above as being missing from Tables 3-6 and 3-7, as well as Danesh Yazdi M, Wang Y, Di Q, Wei Y, Requia WJ, Shi L, Sabath MB, Dominici F, Coull BA, Evans JS, Koutrakis P, Schwartz JD. Long-Term Association of Air Pollution and Hospital Admissions Among Medicare Participants Using a Doubly Robust Additive Model. Circulation, 2021. DOI: 10.1161/CIRCULATIONAHA.120.050252

The second study has an analysis restricted to PM2.5 concentrations below 10 μ g/m³ with a median value of 8 μ g/m³ in that subgroup.

P3-115. The Table omits many studies with restricted analyses. The include the Danesh Yazdi study above as well as:

Danesh Yazdi M, Wang Y, Di Q, Requia WJ, Wei Y, Shi L, Sabath MB, Dominici F, Coull. BA, Evans JS, Koutrakis P, Schwartz. JD. Long-term Effect of Exposure to Lower Concentrations of Air Pollution on Mortality among Medicare Participants and Vulnerable Subgroups. Lancet Planetary Health, 2021, **5**(10): e689-e697

Wei, Y., M. Yazdi, Q. Di, W. Requia, F. Dominici, A. Zanobetti and J. Schwartz (2021). "Emulating causal dose-response relations between air pollutants and mortality in the Medicare population." Environmental Health 20 Article number: 53 (2021).

P 3-119, Table 3-11 omits the Wei study cited immediately above this line, as well as Schwartz JD, Yitshak-Sade M, Zanobetti A, Di Q, Requia WJ, Dominick F, Mittleman MA. A self-controlled approach to survival analysis, with application to air pollution and mortality. Environment International, 2021, 157; 106861.

P 3-126 "Thus, uncertainty in hybrid model predictions becomes an increasingly important consideration as lower predicted concentrations are considered." As shown previously this statement is false. The uncertainty in the hybrid model is lowest below 10 μ g/m³, and decreases further at e.g. 5 μ g/m³. This is not surprising. The large majority of the exposures and measurements of PM_{2.5} in the U.S. have been below the NAAQS for almost two decades. That is where the bulk of the measurements are, and there is no reason to believe that modeled exposures have greater error in the exposure range where there is more data available to fit the models than in the exposure range where the models rely on less data.

P 3-128 "There is less certainty in the shape of the concentration-response curve at mean annual PM2.5 concentrations generally below 8 μ g/m³," Again, this is not true, and the statement seems to derive from the PA only attributing demonstrations of effects below 8 μ g/m³ to studies with means below 8 μ g/m³. This is illustrated in the below figure from Wei 2021. With over 60 million person-years in each decile of exposure it provides considerable confidence that effects continue below 8, as shown by the confidence intervals in the figure.

