The authors experienced a 38% loss of sample size among those patients using manual diaries. Accounting for nonresponse bias almost certainly would have made these figures worse.[48]

In its response, EPA says only that it "recognizes that PEF measurements have been shown to be more variable than FEV_1 in some studies" (U.S. Environmental Protection Agency 2008e, p. 48).[49] The relative variability of PEF to FEV_1 measurements is a *non sequitur*, but it turns out to be a revealing one nevertheless. We deal with this in the following subsection.

In their study of asthmatic adults, Ross et al. (2002) acknowledge that they had problems with data quality – problems that are inherent to the research design:

> Our study also had shortcomings that are shared by most panel studies, such as the possibility of incorrect data recording by study participants. Previous surveys have reported that diary cards with self-reported PEFR and symptom data may contain a high number of invented or retrospective entries.[50]

EPA appears to have been well aware of the problems posed by diary recordation of pulmonary function data. The lead author of Ross et al. (2002) is an employee of EPA's Office of Air Quality Planning and Standards.

(iv) Information quality defects associated with inter-maneuver variability

One of the two studies EPA cites for the observation that PEF measurements are more variable than FEV_1 is the study comparing alternative devices by Vaughan et al. (1989) – a study with which we previously had been

---

[48] Electronic data collection assures that the data collected are accurate, but it does not assure that data will be collected. Medical researchers have concluded that both electronic data collection and sufficient motivation to adhere to the prescribed data collection regimen are essential. See Reddel et al. (2002).

[49] See (U.S. Environmental Protection Agency 2006a, pp. 7-27 to 27-47). EPA also tries to rebut Kamps with a paper by Lippmann and Spektor; part of the appeal may be that Lippmann is a longtime CASAC member. The rebuttal paper is off target; it is a comparison of the performance of alternative devices and has nothing to do with the reliability of data recorded in diaries.

[50] See Ross et al. (2002, p. 577, internal citations omitted). They authors add: "We would, however, expect these limitations to bias the study results in the direction of nonsignificance." They do explain why this would be so.

unfamiliar. They did more than just compare inter-instrument variability, however. They also estimated inter-maneuver standard deviations for three maneuver trials. The mean inter-maneuver standard deviations across 102 patients was 3.01% (FEV_1), 5.12% (mini-Wright peak flow meter), and 7.2% (Jones Pulmonar Spirometer). Mean inter-maneuver standard deviations were higher for patients with COPD (3.11%, 5.50%, and 7.43%) than for patients with no disease (2.82%, 4.29%, and 7.03%) (p. 560).

Several of the epidemiological studies on which EPA staff use for risk assessment rely on pulmonary function measurements. Some cite the ATS guidelines as the basis for their procedures, but at least one crucial study relied on caregivers to do this (Mortimer et al. 2002). If the ATS guidelines are followed, then researchers will have obtained between three and eight separate measurements and they will have made crucial decisions concerning which single value is most representative of the subject's contemporaneous pulmonary function. Choosing any single value, taking an average or performing some other calculation, and discarding the remaining data all create a false sense of precision. Typically, epidemiologists use the difference between subjects' pre- and post-exposure performance as their measure of effect due to exposure. Subtracting a pair of fixed values, when each is actually uncertain, exacerbates the excess precision problem.

We took a closer look at these studies and discovered that inter-maneuver variability is never accounted for. In every case, a single value is recorded as representative for each test, often with a very high degree of implied precision. Similarly, differences between pre- and post-exposure pulmonary function are calculated to retain this precision. Instead of taking account of uncertainty and variability inherent to the pulmonary function test, both are discarded. Thus, all reported standard errors in these observational studies are significantly underestimated. Odds ratios and relative risk measures that are reported to be statistically significant almost certainly are not.

Korrick et al. (1998) is representative. They obtained expiratory flow measures from hikers at Mt. Washington, New Hampshire, following the ATS guidelines issued in 1987:

> Each participant performed a <u>minimum of three and a maximum of eight</u> forced expiratory maneuvers before the day's hike and again after returning to the base. For each hiker, mean values for forced expiratory volume in 1 sec (FEV_1) and forced vital capacity (FVC) were the <u>means of the two or three best acceptable and reproducible (± 5%) values</u>.

The adjusted mean percentage changes in FEV_1 and FVC reported by Korrick et al. (1998) were 5.1% and 4.3%, respectively. These figures are about the half the

magnitude as the inter-maneuver standard deviations reported by Vaughan et al. (1989). Had Korrick et al. (1998) taken account of inter-maneuver variability (e.g., by recording the value for each "acceptable" maneuver), it is very unlikely that the effects they reported would have been statistically significant. The PEFR decrements reported by Mortimer et al. (2002) – less than 1% -- are one-ninth to one-twelfth of the mean inter-maneuver standard deviation for PEFR tests reported by Vaughan et al. (1989).

Vaughan et al. (1989) is 19 years old. EPA staff have long been aware that pulmonary function test measurements are not fixed, but highly variable. They have chosen not to include this important information in their discussion and analysis of the short-term epidemiological studies that show weak but barely statistically significant evidence of respiratory effects from ozone exposure below the 1997 NAAQS. Although EPA cites Vaughan et al. (1989) in its Response to Comments, it does not list the paper as a reference; the paper is discussed in the Criteria Document only with respect to variance in PEFR measurements (U.S. Environmental Protection Agency 2006a, p. 7-29); and it is missing entirely from the Staff Paper. In the Criteria Document, EPA staff summarize and discuss many short-term epidemiological studies in which small differences in pulmonary function are estimated and determined to be statistically significant. Not once does EPA staff mention that inter-maneuver variance exists, much less than it had been routinely discarded.

When epidemiologists try to use crude clinical diagnostic tools for sophisticated research purposes like estimating very weak associations, the consequences of discarded inter-maneuver uncertainty and variability become quite serious. Epidemiologists have achieved marginal statistical significance by employing innovative techniques (e.g., distributed lag models) and made expansive claims about the policy relevance of their work. Had they accounted for inter-maneuver variance instead of discarding it, however, the statistical significance of these weak associations would have vanished.

(v) Information quality defects resulting from nonresponse bias

We have already noted that some of the studies on which EPA relies have response rates too low to reasonably assume that nonresponse bias is not a problem. The response rate in Mortimer et al. (2002) was no greater than 60%. Korrick et al. (1998) used a convenience sample and achieved a 78% response rate.

Recent guidance issued by the Office of Management and Budget codified in writing the longstanding but informal government-wide statistical policy which requires that surveys with response rates below 80% include a rigorous

nonresponse bias analysis in order to qualify for approval under the Paperwork Reduction Act (Office of Management and Budget 2006). Both Mortimer et al. (2002) and Korrick et al. (1998) would have failed this test, and possibly also because they obtained a convenience sample. In the study by Gent et al. (2003), 357 children were determined to be eligible, 75 (21%) refused or were lost to follow-up, and 14 (4%) withdrew, leaving a response rate of 76%. The authors say nothing about any efforts they made to estimate nonresponse bias, and it is assumed but not shown that their original sample was representative. They simply assume representativeness sufficient to justify the statistical tests they perform and assume away nonresponse bias.[51]

(vi) EPA's use of PEFR data depends on whether the results support an inference of pollutant-related health effect

In the scientific record for the ozone NAAQS review, EPA considers pulmonary function test data to be valid and reliable despite the problems discussed in the previous five sections. In its Response to Comments, EPA persists in defending the use of "small inexpensive flow meters" apparently because a longstanding CASAC member likes them.[52] In its discussions in the Criteria Document, Staff Paper, and elsewhere, there is no hint of doubt that pulmonary function measurements are anything but reliable.

On August 30, 2007, about six weeks after finalizing these documents and publishing the proposed rule, the Agency separately distributed for public comment and CASAC review its draft Integrated Science Assessment for nitrogen oxides (U.S. Environmental Protection Agency 2007b). Unsurprisingly, some of the same studies that are relevant to estimating human health risks from ozone also are relevant to estimating analogous risks from NOx. Very surprisingly, however, in the NO_x ISA EPA says that pulmonary function test data are "notoriously" unreliable:

---

[51] Korrick et al. (1998) claim that their convenience sample was representative because "[t]he study researcher and hikers were unaware of the ambient O_3 or other pollutant levels." Lack of awareness of ambient ozone levels helps avoid strategic behavior but it cannot achieve sample representatives. Moreover, the purpose of the study was communicated to prospective subjects, and it would not be a surprise if some hikers tried to "help" the researchers prove their point.

[52] See EPA (2008d, p. 33, citing a paper co-authored by Lippmann). Lippmann proposed the citation in his comments on the draft Criteria Document (Henderson 2005a, p. C-66), and EPA obliged. However, the issue at hand was not diary reliability but the relationship between FEV_1 and PEF.

> Reliable data are notoriously difficult to come by using portable peak flow measuring devices (p. 3-16).

EPA summarizes – and dismisses – several studies in which pulmonary function data were collected. Among them: the study by Mortimer et al. (2002), the same study of asthmatic children that, in the ozone Staff Paper, EPA said "suggest[s] that O_3 exposure may be associated with clinically significant changes in PEF in asthmatic children" and identified "plausible biological mechanisms that would explain delayed effects consistent with the distributed lag models that yielded that only statistically significant results."

In the ozone Staff Paper, EPA considers the use of PEFR monitors by Mortimer et al. (2002) to be state of the art and their results persuasive:

> The multicities study by Mortimer et al. (2002), which provides an asthmatic population most representative of the United States, and several single-city studies indicate a robust association of O_3 concentrations with respiratory symptoms and increased medication use in asthmatics (U.S. Environmental Protection Agency 2007m, p. 3-11)

In the NO_x Integrated Science Analysis, however, their work was no good at all.

These differences may be extreme but they are not random. The difference in EPA staff treatment of Mortimer et al. (2002) in the ozone and NO_x cases cannot be the result of a change of heart about pulmonary function tests. The only material difference is that Mortimer and coworkers found statistically significant effects for ozone but no effects for NO_x. Consistent with the Envelope Theory we enunciated in Section I.C, Mortimer et al. (2002) pushes the ozone risk envelope outward (and thus it is valid and reliable) but pushes the NOx risk envelope inward (and thus it must be discarded).[53]

This phenomenon is not an isolated occurrence. In its Response to Comments, EPA is dismissive of the randomized panel study of asthmatic children by Schildcrout et al. (2006) (U.S. Environmental Protection Agency

---

[53] Discarding Mortimer et al. (2002) did not pose much of a barrier to the NO_x health risk characterization: EPA staff found other studies to support its predictable conclusion that NO_2 posed a health risk to asthmatic children:

> Taken together, these studies indicate that short-term exposure to NO2 is associated with respiratory symptoms in children…. For children, the results of new multicity studies provide substantial support for associations with respiratory symptoms, particularly in asthmatic children (U.S. Environmental Protection Agency 2007b, p. 3-31).

2008e, p. A-3 to A-5). EPA faulted it for having just 990 subjects. "As a result," EPA writes, "the total number of children observed by Schildcrout et al. is not comparable to other large multi-city studies that examined the effect of O_3 concentrations on asthma exacerbation, such as Mortimer et al. (2002)." This is an especially odd complaint, inasmuch as the study by Mortimer et al. (2002) included 846 children.[54]

EPA's low opinion of Schildcrout et al. (2006) is limited to ozone, however. In EPA's final Integrated Science Assessment for SO_2, EPA says "the strongest epidemiological evidence for an association between respiratory symptoms and exposure to ambient and SO_2 comes from two large multi-city studies" -- Mortimer et al. (2002) and Schildcrout et al. (2006). The difference is that Schildcrout et al. (2006) reported a statistically significant positive association between SO_2 and respiratory symptoms, but no association with ozone. EPA likes Mortimer et al. (2002) for both ozone and SO_2; Mortimer et al. (2002) found positive associations for both.

3. Peer review practices

In our RFC, we raised questions about the peer review practices of scholarly journals and noted how they differed from government peer review. Most importantly in this context, it is EPA policy to fully incorporate information quality into its peer review practice (U.S. Environmental Protection Agency 2006e). Few, if any, scholarly journals have followed suit. Thus, there is no reason to assume that information quality principles play any significant role in journal peer review.

We also raised questions about EPA's Clean Air Scientific Advisory Committee (CASAC) as a peer review body. We noted that its statutory charge included both reviewing EPA's risk assessment and providing policy advice. CASAC's policy advice function confounds its scientific review, making it difficult – and, in some cases, impossible – to discern when it is performing scientific review and when it is delivering policy advice.

---

[54] EPA then resorted to a double negative to reinterpret the authors' no-effect finding, and demand that evidence of no-effect be accompanied by proof, just as we have hypothesized in our Envelope Theory of EPA Staff Ozone Risk Characterization: "Although Schildcrout et al. did not find an association between O_3 concentrations and asthma exacerbation, Shildcrout does not imply the results are inconsistent with those previously found because a thorough evaluation of study populations, uncertainty in parameter estimates, precise scientific questions, and additional comparisons between studies that examined the effect of O_3 exposure on asthma exacerbations has not been conducted." See EPA (2008e, p. A-5, emphasis added).

This problem could have been significantly reduced if EPA had included information quality principles within its charge to CASAC. It did not. EPA's Information Quality Guidelines are not even mentioned in the charge, and unsurprisingly, none of CASAC's reports has anything to say about the subject. EPA may have established a policy whereby information quality is incorporated in peer review, but at least with respect to the CASAC process, that policy has yet to be implemented.

In EPA's Response to Comments, the Agency is silent with respect to these issues.

B. Non-disclosure of critical studies and analyses

In our RFC we said that EPA "excluded scientific information for reasons other than defects in information quality" relevant to determining Policy Relevant Background (PRB) (p. 62.). We highlighted the data reported by Vingarzan (2004), Oltmans et al. (2006), and Brown (2007a).

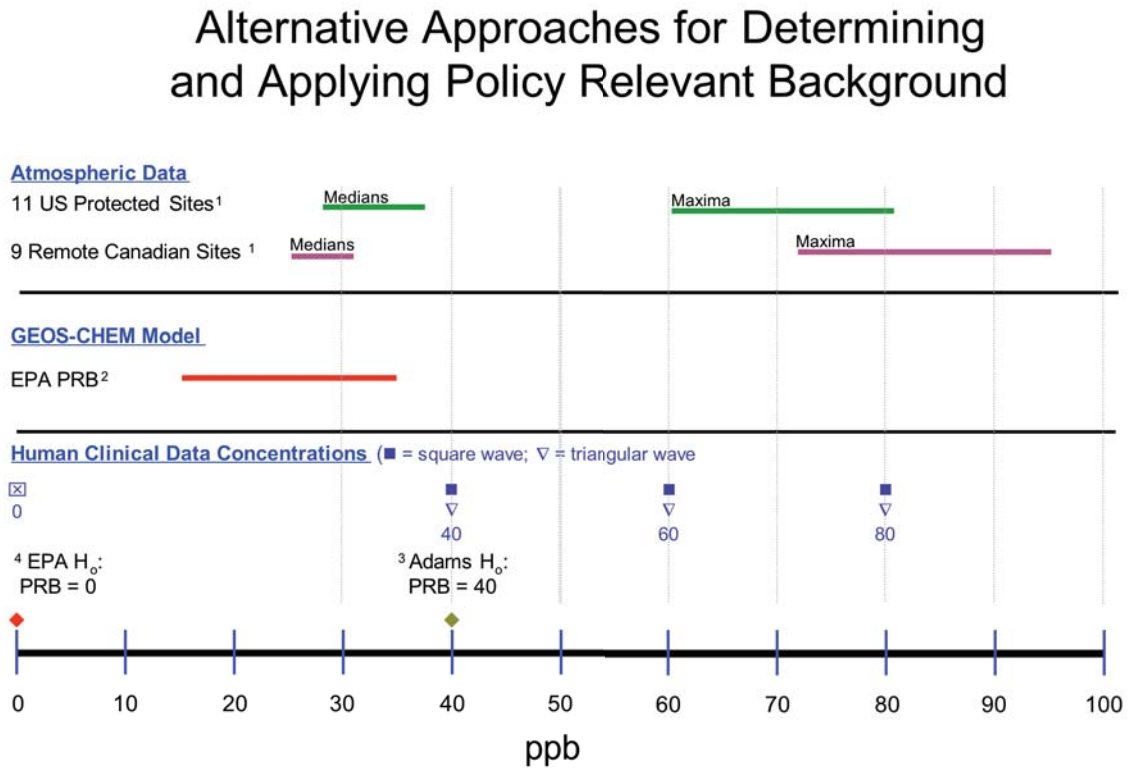
1. Vingarzan (2004) and Oltmans et al. (2006)

In its response, EPA says that the study by Vingarzan (2004) "was considered by EPA" (U.S. Environmental Protection Agency 2008d, p. 96), but as we have noted in many earlier contexts, the nature of that consideration is not at all clear. Vingarzan's data are summarized in the Criteria Document (U.S. Environmental Protection Agency 2006a, pp. 10-49 to 10-50), but that appears to be the sum total of EPA's "consideration." Subsequently, these data disappeared into a black hole: they are not referenced, much less "considered," in the Staff Paper.

Comparing Vingarzan's data as reported in the Criteria Document against the range EPA selected (15 to 35 ppb) shows the bias we allege exists in EPA's determination of PRB:

- At background stations in protected areas of the U.S. (Table AX3-11), the means (standard errors) of 11 lower- and upper-bound medians are 28 (2.8) ppb and 37 (4.6) ppb, respectively. EPA's lower-bound PRB (15 ppb) is 4.6 standard errors below the mean of the lower bounds. EPA's upper-bound PRB is 0.4 standard errors below the mean of the upper bounds.
- Concentrations at background stations in Canada (Table AX3-12), the means (standard errors) of nine lower- and upper-bound medians are 26 (0.5) ppb and 31 (0.7) ppb, respectively. EPA's lower-bound PRB (15 ppb) is 22 standard errors below the mean of the lower bounds. EPA's upper-bound PRB is 5.7 standard errors above the mean of the upper bounds.

The figure below illustrates visually the gap between observational data, EPA's GEOS-CHEM model, and the controlled human data on personal exposure.[55]



Sources: [1]Vingarzan (2004), [2]EPA (2008), [3]Adams (2006), [4]Brown (2007)

EPA acknowledges that it did not "consider" the paper by Oltmans et al. (2006). This EPA attributes to the study having been "published after completion of the Criteria Document" (p. 96). The paper was accepted for publication on

---

[55] The ranges labeled "medians" are arithmetic averages of the lower- and upper-bound annual medians reported by Vingarzan. Similarly, the ranges labeled "maxima" are the arithmetic averages of the reported lower- and upper-bound annual maxima. Averaging reduces the influence of individual annual values.

The concentrations used in Adams also are reported, with the PRBs that Adams and EPA, respectively, assumed in hypothesis tests. Elsewhere we deal with the problem of EPA's transparent effort of ex post data mining. The relevant point here is that the extremism inherent in EPA's use of zero ppb as the "background" level in its hypothesis is visually obvious.

January 18, 2006 (before the "completion" of the Criteria Document) and posted online on March 22, 2006. Oltmans et al., (2006) shows that background ozone at 33 remote locations varies significantly within and across years, and by location. Values similar to EPA's upper-bound for PRB (0.035 ppm) have been frequently observed, but no example is reported in which background levels ever approached EPA's lower-bound (0.015 ppm), even though some values include ozone from anthropogenic sources.

With respect to the substance of this scientific information, EPA's response is just dismissive: "EPA has already discussed the fact that there is spatial and seasonal variability in PRB in the Criteria Document and Staff Paper and the GEOS-CHEM model runs also show this spatial and seasonal variability" (U.S. Environmental Protection Agency 2008e, p. 93). Consistent with the Envelope Theory, scientific studies that push the ozone risk envelope inward -- or, in this case, reduce the potential risk reduction that a more stringent NAAQS might achieve -- will be "considered" and "discussed" before they are discarded.

2. Brown (2007a)

The most obvious example of EPA's gerrymandered scientific database is EPA's own reanalysis of the Adams (2006a) clinical data. EPA placed this document into the docket six days before the Administrator signed the proposed rule (Brown 2007a). This reanalysis is the lynchpin to EPA's scientific database, EPA's denials notwithstanding.[56] It is the only putative scientific basis EPA has for claiming that there is clinical evidence that 0.06 ppm ozone causes any decrement in pulmonary function, adverse or otherwise, in healthy adults. Yet it appeared in the docket at the eleventh hour – without public notice and comment (unlike the Criteria Document) and without peer review (unlike Adams (2006a)).

In our RFC, we characterized EPA's action as a clear violation of applicable information quality guidelines:

> This reanalysis is fully subject to information quality standards and does not benefit from the weak rebuttable presumption of objectivity

[56] "[T]he Brown Memorandum [i.e., Brown (2007a)] is not a crucial element of the staff's policy recommendations, as it was prepared after completion of the Staff Paper, or the Administrator's final decision" (U.S. Environmental Protection Agency 2008d, p. 20). This statement is counterfactual. EPA relied specifically on Brown (2007a) in the preamble to the proposed rule (U.S. Environmental Protection Agency 2007h, p. 37828) and even more so in the preamble to the final rule (U.S. Environmental Protection Agency 2008b, pp. 16454-16455).

> because it has not been peer reviewed. Moreover, because it reaches conclusions opposite of the researcher, it is equivalent to a new study inserted into the record in a discriminatory fashion. It is beyond dispute that EPA would not have accepted a new analysis of the Adams data submitted by a third party on June 14, 2007, unless perhaps it supported the staff's policy recommendations. EPA clearly displays a discriminatory preference for data and analyses that support staff risk management preferences, an obvious information quality defect (National Association of Manufacturers 2007, p. 17).

(a) EPA's reanalysis of Adams (2006a) is technically defective.

In our RFC, we objected on information quality grounds to the technical merits of EPA's reanalysis of the Adams (2006a) data (National Association of Manufacturers 2007, p. 18). This reanalysis (Brown 2007a) consists of post hoc statistical tests of selected data originating in an admittedly low-quality analytic review ("visual comparison" and "cursory evaluation," p. 3) in which EPA staff homed in on pulmonary function responses from two of the 30 subjects. The reanalysis was structured for the purpose of minimizing Type II error (failing to reject the no-effect hypothesis when in fact it is false, pp. 5ff). The reanalysis compares changes in pre- to post-exposure responses for square- and triangular-wave exposures as if the intermediate effects during the 6.6 hour test period are unimportant. In the analysis, EPA provided no bona fide external technical defense for the statistical methods it used, and after being challenged, it has failed to provide a technical defense in its Response to Comments. In both instances, EPA cites itself as its technical authority.

(b) EPA's explanation for why it reanalyzed selected data from Adams (2006a) is materially incomplete and misleading

In its Response to Comments, EPA says that its reanalysis was merely "a logical progression" from Adams (2006a) that was somewhat delayed only because Adams' papers were "not published until 2006" (U.S. Environmental Protection Agency 2008e, p. 20). The logical progression we see is less scientific than policy-driven. With the help of Adams,[57] we have reconstructed the timeline of events to prove that EPA's explanation is highly misleading, and thus in its Response to Comments EPA has committed a new violation of the Agency's information quality guidelines.

---

[57] William C. Adams, personal communications during July and August 2008.

Contrary to EPA's claims, EPA staff began to reanalyze Adams (2006a) in search of statistically significant effects at 0.06 ppm about 18 months before Brown (2007a) was placed in the docket. The key event appears to be a request from CASAC that Adams (2006a) be included in the Criteria Document. That occurred in December 2005. From that date onward, EPA staff obtained portions of Adams' data, reanalyzed them, presented their results informally at an EPA-sponsored symposium, and tried to persuade Adams to join them in supporting their statistical reanalysis.

To show how misleading and self-serving is EPA's version of the story, the **facts reported by EPA in its Response to Comments are presented in bold green font** and ***EPA's significant omissions are presented in bold italic red font***:

- September 13, 2002: EPA hires Adams as a consultant to co-author Chapter 7 of the Criteria Document. EPA asks Adams to update the summary of human ozone exposure research and instructs him to ignore all studies not accepted for publication; his review excludes the recent research that became Adams (2006a, 2006b) which had not been submitted.
- August 2005: Adams submits his updates of human ozone exposure research following CASAC review of the first draft Criteria Document. It excludes any discussion of the research that became Adams (2006a, 2006b). These studies are subsequently included in the Criteria Document, but EPA never asks Adams to revise this update, nor does EPA ask Adams to review the summary they write.
- July 28, 2005: Adams (2006a) is accepted for publication in the refereed journal *Inhalation Toxicology*. Adams confirms results for 0.08 ppm obtained in Adams (2003) but reports no statistically significant effects at the previously untested concentrations of 0.06 ppm and 0.04 ppm.
- November 2, 2005: Adams (2006b) is accepted for publication in the refereed journal *Inhalation Toxicology*. Adams largely confirmed results obtained by (Hazucha et al. 1992).
- December 9, 2005: ***EPA's James Brown notifies Adams that CASAC panel member Henry Gong has requested a copy of Adams (2006a). Adams sends Brown a copy of the corrected galley proofs of Adams (2006a). The galleys contain only a few minor handwritten corrections.***
- December 15, 2005: ***Brown notifies Adams that CASAC has asked EPA to include Adams (2006a) in the Criteria Document.***

- December 21, 2005: *EPA's Harvey Richmond requests from the American Petroleum Institute (API), at a minimum, the pre- and post-exposure (but not hourly) data from Adams (2006a).*
- January 10, 2006: *Richmond requests from Adams the pre- and post-exposure (but not hourly) FEV_1 data in Adams (2002, 2003, 2006a). This is the first and only time Adams received any request for data from EPA.*
- January 17, 2006: *API public comment on the initial draft Staff Paper asks EPA to include Adams (2006a) in the final risk assessment and notes that EPA staff have asked Adams for a portion of the data from Adams (2006a).* **EPA accedes to API's request that Adams (2006a) be included,** *but does not acknowledge CASAC's identical prior oral request.*
- January 20, 2006: *Adams sends pre- and post-exposure FEV_1 (but not hourly) data in Adams (2006a) to Richmond.*
- January 23, 2006: *API responds to Richmond noting their earlier request to Adams to consider providing the data; Richmond confirms Adams' provision of the requested data.*
- January 23, 2006: *Adams sends pre- and post-exposure FEV_1 (but not hourly) data in Adams (2003) to Richmond.*
- January 25, 2006: *Adams sends pre- and post-exposure FEV_1 (but not hourly) data in Adams (Adams 2002) to Richmond.*
- February 10, 2006: *In its letter to the Administrator on the second external review draft of the Criteria Document (Henderson 2006a, p. 5), CASAC specifically recommends that Adams (2006a) be included.*
- December 14, 2006: *Brown presents summary results of his reanalysis of Adams (2006a) at "EPA Workshop on Interpretation of Epidemiologic Studies of Multipollutant Exposures and Health Effects," Chapel Hill, N.C. (Brown 2006).*
- January 3, 2007: *Brown provides Adams with his draft reanalysis (Brown 2007b) and seeks Adams' collaboration on a final version.* The draft text includes a courtesy acknowledgement that Adams provided the data that EPA had requested.
- January 6, 2007: *EPA adds to the draft final Staff Paper a summary of Adams (2006a), including material "not mentioned in the CD," and a summary of Brown's reanalysis (U.S. Environmental Protection Agency 2007k, pp. 3-5 to 3-9). Brown (2007b) is neither cited nor*

disclosed, thereby limiting public comment and CASAC review to the nonreproducible summary presented in the draft final Staff Paper.

- January 9, 2007: Adams declines Brown's offer to collaborate on Brown's interpretation of the FEV_1 results for 0.06 ppm from Adams (2006a). Adams tells Brown that he disagrees with Brown's focus on pre- and post-exposure FEV_1 data only and Brown's choice of statistical methods.
- January 24, 2007: Richmond requests from API a copy of Adams' 1998 draft final report to API cited in Adams (2006a, p. 133). The 0.12 ppm exposure results, but not the 0.06 ppm results, are published in Adams (Adams 2000). API makes no response to Richmond's request; this is reported by EPA as **"API refus[es] to provide Dr. Adams technical report describing that data."**
- March 4, 2007: **During the CASAC teleconference to review the final Staff Paper, presentations are made by Adams (2007) and Richard Smith (University of North Carolina--Chapel Hill) (Smith 2007b)**:
  - ***Adams objects to several aspects of EPA's summary of his work in the draft final Staff Paper, which he believes is not fairly or accurately presented.***[58]
  - ***Smith "used the same statistical approach" that Brown used in his reanalysis of the Adams data.***[59] ***He "also utilized t tests to evaluate the statistical significance of the Adams data…"***[60] ***Smith "specifically indicated that the FEV_1 responses, in the Adams (2006[a]) study, following the two 0.06 ppm O_3 exposures were statistically different from the FEV_1 responses following filtered***

[58] See Adams (2007). Adams raised three specific objections: (1) EPA's use of standard errors instead of standard deviations, which he says reduces subject variability by a factor of about 5.5, thereby making apparent statistical significance much easier to observe; (2) EPA's statement that exposure to 0.06 ppm causes small group mean FEV1 decrements, which Adams says were not statistically significant; and (3) EPA's claim that the fraction of Adams' subjects who experienced greater than 15% FEV1 decrements was lower than in EPA chamber studies because of adaptation to higher ozone levels in Davis CA than in Chapel Hill NC, which Adams says is simply factually incorrect based on EPA's own ambient monitoring data.

[59] See EPA (2008e, p. 22).

[60] See EPA (2008e, p. 27).

air exposures using a paired t test."[61] Smith's purpose in using the "same statistical approach" was to reproduce results reported by EPA in the draft final Staff Paper (U.S. Environmental Protection Agency 2007k, pp. 3-5 to 3-9), not to provide guidance on correct statistical procedures.

- ***Smith says the draft final Staff Report understates the confidence interval on its tests of the proportion of individuals who showed an FEV_1 decrement greater than 10%. Smith calculates the confidence interval as 0.8% to 22%, not 6% to 16%.***
- ***Smith objects to EPA's use of filtered air as the baseline from which to measure the effects of 0.06 ppm. "In making policy-relevant comparisons, those with 0.04 ppm ozone level are more relevant than those with filtered air, which does not represent a realistic background level."***
- ***Smith objects to EPA's use of a logistic response curve because it "assumes that the response curve fitted to higher ozone levels can be extrapolated downwards to 0.06 ppm. Given the large uncertainty in the probability of response at 0.06 ppm ozone, I do not believe the staff paper's conclusions on this point are justified."***
- ***Smith says, "[W]hen all possible comparisons are taken into account, there is insufficient evidence to conclude that there is any well-defined response to ozone exposure below the 0.080 ppm level."***
- ***Smith recommends against using the statistical method employed by EPA.***

- March 8, 2007: Richmond makes second request to API for a copy of Adams' 1998 draft final report to API.
- March 15, 2007: API declines Richmond's March 8th request and encourages EPA to rely on published 0.06 ppm exposure results (Adams 2002, p. 741).
- March 16, 2007: Richmond asks API for additional details about the 0.06 ppm exposure cohort described in Adams (Adams 2002, p. 741; 2006a, p. 133).

[61] See EPA (2008e, p. 21).

- June 20, 2007: **EPA places its reanalysis of the Adams data (Brown 2007a) in the rulemaking docket. The text says EPA "obtained" the data from Adams (2006a)** ***but does not say that Adams' provided these data in January 2006. The text implies that EPA performed its reanalysis in response to March 2007 comments to CASAC by Smith (2007b), and says nothing about the December 2006 presentation in which preliminary results of EPA's reanalysis were disclosed (Brown 2006) or the January 2007 draft of the reanalysis (Brown 2007b), both of which predate Smith (2007b).***[62]
- March 15, 2008: In its Response to Comments:
  - **EPA says API asked that Adams (2006a) be included in the Criteria Document*****, but does not acknowledge that CASAC had already made the same request.***
  - **EPA says Smith's public comment to CASAC initiated EPA's reanalysis of the data in Adams (2006a)*****, but does not acknowledge that:***
    - ***EPA began its reanalysis in early 2006***
    - ***Summary results from EPA's reanalysis were first disclosed by EPA in December 2006***
    - ***Brown shared a draft of his reanalysis with Adams in January 2006 seeking Adams' collaboration.***

The most reasonable interpretation of this history is that EPA staff fully intended not to include Adams (2006a) in the Criteria Document because it did not show statistically significant effects at 0.06 ppm. However, once CASAC asked that it be included, EPA staff had to find a way to discredit Adams' conclusions without challenging Adams' professional reputation). EPA staff resolved to obtain and utilize only selected parts of Adams' dataset, thus making the statistical challenge of "finding" significant effects less daunting. Brown

---

[62] In its Response to Comments (U.S. Environmental Protection Agency 2008e), EPA several times refers to Smith's March 2007 comments to CASAC, which are referenced herein as RL Smith (2007b). However, the list of references cites only Smith's October 2007 public comment to EPA, which is referenced herein as RL Smith (2007a). In addition to selectively presenting RL Smith (2007b) and materially misrepresenting them as supporting EPA's statistical methods, EPA never responds to any part of RL Smith (2007a) even though Smith says that Brown materially misrepresented his work in Brown (2007a). This new information quality error persists throughout EPA's Response to Comments.

telegraphed their preliminary success to the public via his December 2006 presentation at the EPA-sponsored conference in North Carolina. Subsequently, Brown offered Adams an opportunity to assist in the reanalysis -- the most conventional of ways used to co-opt academic researchers – but Adams refused.

The prospect that EPA would reinterpret a published, peer-reviewed study to reach an opposite conclusion alarmed the study's sponsor, the American Petroleum Institute. API hired Smith to replicate EPA's alternative analysis. Smith was mostly able to do so, but discovered significant technical errors in Brown's work and further opined that Brown's entire approach was fundamentally flawed because it failed to account at all for multiple comparisons. However, in the process of describing fully why EPA ought not to proceed along the course it set forth in the Staff Paper, Smith unwittingly gave EPA staff a road map for how to correctly perform the calculations for its preferred (but fundamentally flawed) statistical method. EPA staff seized the opportunity to avoid technical error and recast the reanalysis as a "confirmation" of Smith's work. EPA placed the final reanalysis in the rulemaking docket the same day the Administrator signed the proposed rule.

There is no record that any CASAC member actually focused on the issues raised by Adams and Smith during the March 2007 teleconference. Brown's work was never subjected to review by CASAC because it was placed in the docket after CASAC review was completed. Brown asserted that unnamed CASAC members "supported" his statistical approach,[63] but this cannot be documented because EPA's Science Advisory Board does not make transcripts of teleconferences.

We have looked elsewhere for evidence of CASAC support for the EPA staff's statistical methods. In comments prepared in August 2006 – four months before Brown disclosed results from his preliminary reanalysis (Brown 2006) and 10 months before EPA placed his work in the rulemaking docket (Brown 2007a) -- CASAC panel member Frederick Miller supported highlighting selected

[63] Brown (2007a, p. 5): "On the March 5, 2007 teleconference, members of the CASAC O_3 Panel noted the very conservative nature of the statistical test used by Adams to evaluate the research questions posed by the author. These same CASAC Panel members also supported the approach adopted in the OAQPS Staff Paper to evaluate the statistical significance of O_3-related lung function responses associated with pre- versus postexposure responses. The CASAC Panel members also supported the use of the paired t test approach as the preferred method for analyzing the pre- minus postexposure lung function responses."

subjects from the Adams' dataset,[64] but this is the only published suggestion of support we can find.[65] EPA repeats this undocumented claim as fact in its Final Rule[66] and in its Response to Comments.[67] There is no public evidence of any CASAC debate about the propriety of paired *t* tests for analyzing controlled human exposure data in any of the CASAC reports or in the transcripts of CASAC meetings. EPA's assertion is not supported by any factual record, without which it must be inferred that the Agency has responded to our allegations of information quality error by committing additional information quality errors that it hoped would not be detected.

(c) Finding statistically significant effects at 0.06 ppm required EPA to use creative statistical methods

Brown does not say that Adams' choice of hypothesis tests was incorrect, nor does Brown claim that Adams' concern about controlling for multiple comparisons was misplaced. Rather, Brown says he used Adams' data for a completely different purpose than the one for which the study was intended, and therefore multiple comparisons adjustments are not necessary for the reanalysis:

---

[64] "While the discussion of the low level exposures used in the controlled human studies by Adams and colleagues is technically correct that no statistically significant changes were found in FEV_1 compared to filtered air, the fact that a reasonable percent of the subjects had large decrements is glossed over" (Henderson 2006c, p. D-39)

[65] Fellow CASAC panel member Svere Vedal seems to have strongly opposed EPA's cherry-picking of the data. See subsection III.B.2(f) below.

[66] EPA (2008b, p. 16456): "[M]embers of the CASAC Panel noted on the March 5, 2007 teleconference the very conservative nature of the approach used by Adams to evaluate the research questions posed by the author. These same CASAC Panel members also supported the use of the statistical approach (i.e., paired-t test) used in the analysis prepared by the public commenter, which was the same approach later used in EPA's reanalysis, as the preferred method for analyzing the pre-minus post-exposure lung function responses reported in this study."

[67] EPA (2008e, p. 21): "[I]n the Staff Paper, it was noted that a statistically significant difference in FEV1 responses was suggested by a lack of overlap in the standard error of the responses following 6.6 hours of exposure to 0.06 ppm ozone versus filtered air. That interpretation of the data was supported by CASAC review." Elsewhere in the Response to Comments, EPA offers the much weaker defense that its work "was reviewed by the CASAC O_3 Panel and there were no objections expressed" (p. 98). In short, EPA's position is that panel member Vedal's concerns objections (see subsection III.B.2(f) below) do not constitute "objections," and the absence of strenuous peer reviewer opposition is equivalent to peer reviewer endorsement.

> [A]lthough appropriate for the design and intent of the Adams' studies, the multiple comparison correction is overly conservative (increased Type II error and decreased power) for the evaluation of pre- to postexposure changes in FEV_1 between an air and an O_3 exposure and we adopted the standard approach used by other researchers (e.g., Hazucha et al., 1992; Horstman et al., 1995; McDonnell et al., 1991).[68]

The "standard approach used by other researchers" is an example of the logical fallacy known as *argumentum ad verecundiam* -- an appeal to external authority without regard for the truthfulness of the claim itself. EPA's reliance on this logical fallacy constitutes indisputable information quality error.[69]

It also turns out that the "other researchers" cited by Brown do not in fact support his peculiar statistical methods. Horstman et al. (1990, p. 1160) used paired *t* tests to "determine[] the time point at which significant decrements in FEV_1 were observed" during intermediate points of a protocol involving five hours' exposure to 0.00, 0.08, 0.10, and 0.12 ppm ozone. However, they acknowledged that multivariate analysis of variance (MANOVA) for repeated measures would have been "more appropriate." They did not use MANOVA because "this analysis revealed no significant differences ($p = 0.6$) among the four concentrations" – including 0.12 ppm.[70] McDonnell et al. (1991) used paired *t* tests, but a single concentration for a single time period was tested in the study, making the multiple comparisons question irrelevant. Both Hazucha et al. (1992) and Horstman et al. (1995) conducted studies in which multiple comparisons were being made. Hazucha et al. (1992) used two-way ANOVA followed by an

---

[68] See Brown (2007a, p. 5, emphasis added). In our RFC, we cited this statement when we said "Agency staff used the Adams data for purposes that were never intended by the study design" (National Association of Manufacturers 2007, p. 18). In its Response to Comments, EPA recites our information quality objection but provides a reply that it unresponsive (U.S. Environmental Protection Agency 2008e, pp. 21-22).

[69] In fact, the "other researchers" cited in Brown (2007a, p. 5) are not independent. Six of eight co-authors of McDonnell et al. (1991) were at the time EPA employees; one of the non-EPA employees subsequently joined EPA. Of the three co-authors of Hazucha et al. (1992), two were EPA employees. Of the five co-authors of Horstman et al. (1995), the lead and one other co-author was an EPA employee; one of the non-EPA co-authors subsequently joined EPA. The identity of the non-EPA co-author who subsequently joined EPA is James Brown.

[70] The authors say MANOVA was "strongly biased toward a negative outcome" because of limited degrees of freedom, but they do not mean "bias" in a statistical sense. Rather, they mean MANOVA was too demanding as a statistical tool.

unspecified multiple comparisons procedure. Horstman et al. (1995) also used ANOVA, but did not follow with adjustments for multiple comparisons. Brown was a co-author of this study. Thus, Brown's *argumentum ad verecundiam* is worse than merely an appeal to external authority; it's a circular reference to his own prior work.[71]

EPA staff assert that it is acceptable practice to perform simple paired *t* tests on selected results and discard the other data. In our RFC, we asked EPA to disclose an external, independent authority for this statistical method:

> It is inappropriate to obtain a sample, subject its members to a well-designed test, learn that the sample does not yield hoped-for outcome, and in response, abandon the sample in favor of focusing on selected individuals within it. If EPA can find a reputable statistical authority for this procedure, the agency should make its identity known (National Association of Manufacturers 2007, p. 18, emphasis added).

In its response, EPA does not provide a supporting external statistical authority. Smith is a recognized external statistical authority, and he submitted public comments on the proposed rule (Smith 2007a) that are highly critical of EPA's statistical practice, including EPA's mischaracterization and misuse of his public comment to CASAC (Smith 2007b) by Brown (2007a). Instead of responding to the substance of Smith's objections, EPA implicitly suggests that his review is biased because it was funded by API.[72] Where Brown (2007a) engages in the fallacy of *argumentum ad verecundiam* -- appealing to authority instead of logic or fact -- in its Response to Comments EPA commits the highly analogous fallacy known as *argumentum ad hominem circumstantiae* – rejecting claims based on unrelated circumstantial aspects of the opponent – in this case, the opposing

---

[71] The literature that Brown doesn't cite also isn't helpful to his cause. For example, the first controlled human exposure study that tested prolonged exposures – Folinsbee et al. (1988) – used "[m]ultivariate analysis of variance methods appropriate for designs with repeated measurements." Unlike Brown, Folinsbee and his two EPA colleagues refrained from drawing confirmatory inferences based on statistical tests of exploratory hypotheses: "All other tests of hypotheses were of secondary importance and were done only to describe other potential ozone effects and clarify patterns in the data" (p. 30).

[72] "The Brown Memorandum confirms analyses completed by Dr. Smith who was funded by API to perform his analyses and to provide comments to CASAC" (U.S. Environmental Protection Agency 2008e, p. 22, emphasis added). In his public comments, Smith notes that in addition to funding from API, he has received funding from EPA and NIH – neither of which EPA mentions.

authority's source of funding -- rather than the merits of the opponent's argument.

In our RFC we said that Brown (2007a) was not reproducible (National Association of Manufacturers 2007, p. 17) and in its Response to Comments EPA counters that in fact it is (U.S. Environmental Protection Agency 2008e, p. 20)(p. 20). We failed to make our concerns as clear as we should have. The issue at hand is not the algebraic calculation of paired *t* tests. Rather, what is missing from Brown (2007a) is any cogent rationale justifying EPA's analytic approach. The simplest explanation is that EPA staff determined that, to support their policy goals, it was necessary to have statistically significant group mean effects at 0.06 ppm, and when Adams (2006a) came up dry EPA staff needed to find a statistical test that would produce the desired results. The task was challenging in part because EPA staff had never before questioned the statistical methods of Adams or any other researcher performing controlled human experiments. The only way to be able to avoid the burden of making multiple comparisons adjustments was to discard all of Adams' intermediate time period data.

We said in our RFC that EPA staff was so wedded to the policy conviction that the primary NAAQS should be set at 0.06 that they did not merely blur line between science and policy, but they obliterated it. EPA's Response to Comments does nothing to contradict us. In its response, EPA simply "rejects" our arguments without offering an iota of evidence supporting its position (U.S. Environmental Protection Agency 2008e, pp. 21-22). EPA's defense consists of (1) noting that the American Petroleum Institute had asked that Adams (2006a) be included in the scientific record, though conveniently neglecting to mention that CASAC had previously made the same request; (2) cherry-picking Adams' data and statistics textbooks to "discover" statistically significant effects; (3) pretending that these methods were in commonplace use by "other researchers"; and (4) misleading the public to believe that the purpose of EPA's reanalysis of these selected data was only to confirm what Agency staff first learned from Smith's public comment to CASAC (Smith 2007b).

EPA's attempt in the Response to Comments Document to hide Brown (2007a) behind Smith (2007b) is obvious:

> Consistent with common practice for comparing pre-and postexposure [sic] responses to test for whether or not an O_3-related effect is significant, Dr. Smith used a conventional paired t test (U.S. Environmental Protection Agency 2008e, p. 3).

Unfortunately for EPA, Smith is unwilling to serve as the Agency's intellectual shield. He says EPA's statistical procedure is invalid:

> The use of paired t tests to determine significant effects, as originally performed in the EPA Staff Paper and subsequently defended in Brown (2007[a]), is invalid without taking account of the "multiple comparisons" issue (Smith 2007a, p. 1).

Brown (2007a) purports to take account of the multiple comparisons problem but struggles to discover a procedure sufficiently weak that statistical significance – the EPA staff's essential public policy goal -- is still achieved. Having first rejected Scheffé as "conservative," he then rejects the less "conservative" Bonferroni correction[73] because it, too, is too demanding. Needing a threshold no smaller that $p < 0.001$, Brown stumbles upon a solution, though its improvisational reverse-engineering nature cannot be disguised:

> By contrast, a critical p-value might more appropriately be 0.05/5 or 0.01 for assessing pre- to postexposure changes in FEV_1 between an air and an O_3 exposure in the Adams (2006) study.[74]

Brown had many other multiple-comparisons adjustment procedures to choose from, but apparently none fit the bill. In his public comment to EPA on the proposed rule, Smith examined several such procedures, including those devised by Scheffé, Tukey, and Dunnett. All yielded the same result:

> Although the Scheffé procedure used by Adams (2006[a]) is arguably too conservative, alternative options are available through the Tukey and Dunnett procedures. These yield similar results to the Scheffé procedure when performed as part of an analysis of variance, and imply that there is no clear evidence of a decrease in lung function at a mean ozone concentration of 0.06 parts per million (ppm), compared with filtered air.[75]

What Smith makes painfully clear is that the EPA staff's choice of post hoc multiple comparisons adjustment was driven by its need to discard data so that they could dispense with analysis of variance, the standard statistical technique used by scholars, who publish in peer reviewed journals. For its part, in its Response to Comments EPA has nothing to say about Smith's analysis and observations; the Agency is obligated by the Clean Air Act only to respond to

---

[73] In this case, the Bonferroni correction for 90 comparisons yields a p threshold of 0.05/90 = 0.000556.

[74] Brown (2007a, p. 5, emphasis added).

[75] Smith (2007a, p. 1).

those public comments it alone judges to be "significant,"[76] a threshold that Smith (2007a) apparently failed to meet.

(d) The policy-relevant background ozone concentration

Adams estimated whether effects at 0.06 ppm were statistically significant when compared to both filtered air and 0.04 ppm, the same level EPA used for background in its 1997 revision of the ozone NAAQS primary standard (U.S. Environmental Protection Agency 1996b, p. 65726). In both cases, Adams found no statistically significant effects when using statistical methods that account for multiple comparisons, as discussed in the previous subsections. However, Adams did report statistically significant "net" responses for 0.080 ppm whether 0.06, 0.04, or filtered air (i.e., zero ppb) was used as the presumptive background. Adams also tested the difference in FEV_1 response between 0.08 and 0.06 ppm and found that the difference was statistically significant. It is for these reasons Adams concluded that 0.04 and 0.06 pm behaved more like background than like 0.08 ppm.[77]

In his March 2007 comment to CASAC, Smith (2007b) also opined that the tested 0.04 ppm concentration was "more relevant than filtered air, which does not represent a realistic background level" (p.1). EPA staff insist that background is well below 0.04 ppm – so much lower, in fact, that zero ppm is a better proxy for background than 0.04 ppm (Brown 2007a, p. 4, footnote 4). Brown dismisses Smith's objection, once again relying not on any independent authority but on a combination of EPA staff wisdom and an EPA staff policy decision (disguised as "science") to push the policy relevant background below 0.04 ppm:

> As discussed below, we and most authors of the controlled human exposure studies believe that the appropriate approach for testing for an O_3-related response is to compare with filtered air to correct for the effect of exertion in clean air. Additionally, as discussed in the O_3 AQCD (EPA, 2006, AX3-131) and in Chapter 2 of the OAQPS Staff Paper, the scientific evidence supports estimates of policy-relevant background that are in the

---

[76] "The promulgated rule shall also be accompanied by a response to each of the *significant* comments, criticisms, and new data submitted in written or oral presentations during the comment period." See Clean Air Act § 307(d)(6)(B), emphasis added.

[77] The peak exposure in Adams' 0.080 ppm triangular exposure was 0.15 ppm – significantly above the current 1-hour NAAQS. Thus, it is highly inappropriate to construe this exposure protocol as approximating actual ambient conditions at the existing NAAQS.

> 0.015 to 0.035 ppm range in the afternoon during the O_3 warm season, rather than the 0.040 ppm level cited by Dr. Smith (Brown 2007a, p. 4[fn 4], emphasis added).

Brown is the sole named author of the memorandum, so his use of first person plural is stilted at best, and none of the "most authors" he has in mind are identified. Presumably, Brown is referring to the coterie of researchers located in the Research Triangle Park area who perform controlled human exposure studies.[78] These researchers either are EPA employees or are funded by EPA grants. Thus, it hardly would be surprising that, if forced to take a position, "most authors" of the controlled human exposure studies would agree with Brown. They are, after all, his EPA colleagues and answer to the same master. Nevertheless, the EPA staff commitment to using zero ppb as a proxy for ambient background is a policy-driven constraint not supported by scientific evidence.[79]

Apparently unwittingly, Adams stated clearly the conundrum that studies of ambient background levels and his results posed for EPA staff: "[H]ealth effects well may be overestimated in the U.S. Environmental Protection Agency (EPA) risk assessment if [filtered air] is used as the background control" (Adams 2006a, p. 135).

(e) The policy-relevant ozone exposure wave pattern

Previous research has shown that triangular-wave exposures cause earlier respiratory effects than square-wave exposures of the same time-weighted average concentration, and EPA agrees that triangular-wave exposures are more realistic.[80] A principal purpose of the study design in Adams (2006a) was to compare effects under both wave forms to determine whether differences across

---

[78] In the reports published from EPA-sponsored controlled human exposure studies, filtered air (i.e., zero ppb) is used as the baseline for comparisons (e.g., Hazucha et al. 1992; Horstman et al. 1995; McDonnell et al. 1991). None of these studies, however, tested concentrations lower than 0.08 ppm.

[79] In its Response to Comments, "EPA rejects NAM's contention that the Brown Memorandum exemplifies any violation of the information quality standard of objectivity" (U.S. Environmental Protection Agency 2008e, p. 21). In the remainder of the text expounding on this "rejection," EPA argues by *non sequitur*: Brown (2007a) is objective because the sponsor of Adams (2006a) asked that Adams (2006a) be included in the Criteria Document.

[80] EPA agrees that triangular exposures "more closely mimic typical ambient O_3 exposure patterns." See EPA (2007m, p. 3-81).

wave forms that had been observed at concentrations 0.08 ppm and higher also would be present at lower concentrations. For three square-wave concentrations (0.08, 0.06, and 0.04 ppm), Adams devised triangular-wave exposure patterns with the same total exposures, but with higher peaks [0.15, 0.09, and 0.05 ppm) and no exposures less than 0.03 ppm. If the pattern of exposure mattered at these lower concentrations, then stronger effects would be observed with triangular-wave than square-wave exposures.

Adams found that FEV_1 decrements and total symptom scores were significantly greater for triangular-wave exposures after 4.6 and 5.6 hours at 0.08 ppm, but not for 6.6 hours.[81] These results were consistent with results Adams had previously obtained at 0.08 ppm (Adams 2003) and Hazucha et al. (1992) had previously obtained at 0.12 ppm (which Adams also confirmed [(Adams 2006b)]. However, Adams did not observe statistically significant differences between wave patterns at 0.06 or 0.04 ppm. In short, the differential effect of wave-pattern that is detectable at 0.08 ppm and greater concentrations is not apparent at 0.06 ppm and below.

(f) CASAC's "support" for Brown (2007a) is technically infeasible and contradicted by the recollections of some of the principals

Brown (2007a) suggests that the issue is moot because CASAC endorsed his statistical approach:

> On the March 5, 2007 teleconference, members of the CASAC O_3 Panel noted the very conservative nature of the statistical test used by Adams to evaluate the research questions posed by the author. These same CASAC Panel members also supported the approach adopted in the OAQPS Staff Paper to evaluate the statistical significance of O_3-related lung function responses associated with pre- versus postexposure responses. The CASAC Panel members also supported the use of the paired t test approach as the preferred method for analyzing the pre- minus postexposure lung function responses (Brown 2007a, p. 5).

The basis for Brown's claim is hardly self-evident. First, CASAC never reviewed Brown's January 2007 draft memorandum (Brown 2007b). It was not placed in the docket (a requirement for transmittal to CASAC), nor was it on the agenda for CASAC's March 5, 2007, teleconference scheduled to review the draft

[81] Adams used the same statistical methods to adjust for multiple comparisons that Brown (2007a) discarded as "too conservative" with respect to Type I error.

final Staff Paper.[82] Indeed, Brown's January 2007 draft is dated after the release of the draft final Staff Paper (December 2006) even though the draft final Staff Paper includes results from Brown's unpublished January 2007 draft. The earliest CASAC could have seen Brown's reanalysis is June 20, 2007, the day Administrator Johnson signed the proposed rule and EPA placed Brown's finished work product into the rulemaking docket. By this date, CASAC's review was over.

Second, the documentary record indicates that CASAC devoted very little time to this statistical controversy. It appears that CASAC was completely unfamiliar with it until about 15 minutes before the March 5, 2007 conference call. That's when they were provided copies of Adams' and Smith's public comments.[83] Indeed, it appears that EPA worked hard to limit CASAC's exposure to the controversy. Adams and Smith were two of 10 public commenters shoehorned into a 30-minute slot.[84] Under these extraordinary conditions, it would have been quite a remarkable feat for CASAC to digest a pair of oral presentations supplemented by written versions supplied 15 minutes before the conference call began, again review EPA's limited and nontransparent presentation in the final Staff Paper, debate the merits of the competing position, and reach a conclusion – all in the space of maybe an hour -- knowing that the Agency's deadline for disseminating the final Staff Paper was only a couple weeks away.

Third, EPA claims in both its Response to Comments and the preamble to the final rule that it was Smith's public comment to CASAC that created the impetus for EPA's reanalysis of Adams' data.[85] If that were so – and the existence

---

[82] The draft final Staff Paper was made available for public comment on December 27, 2006 (71 Fed. Reg. 77742). The CASAC teleconference to review it was announced on February 5, 2007 (72 Fed. Reg. 25289-5290).

[83] Adams' and Smith's written comments were provide to CASAC by email at 12:45 pm. See the transmittal email from Fred Butterworth to CASAC panel members, Docket No. EPA-HQ-OAR-2005-0172-0075.

[84] The meeting agenda is found at Docket No. EPA-HQ-OAR-2005-0172-0084.1. Public commenters were scheduled from 1:30 pm until 2:00 pm.

[85] From the Response to Comments (U.S. Environmental Protection Agency 2008e, p. 22) "[I]t was a public commenter [i.e., Smith (2007b)] that first placed the analysis of FEV1 responses following exposure to 0.06 ppm O_3 versus filtered air in the public rulemaking docket." From the preamble to the final rule (U.S. Environmental Protection Agency 2008b, p. 16455): "EPA notes that its reanalysis of the Adams (2006)

of Brown's January 2007 draft proves beyond any doubt that the claim is false – then CASAC could not have reviewed the matter carefully enough to "support" the EPA staff position.

Fourth, in their public comments to EPA, Adams and Smith both objected to Brown's claim that CASAC had endorsed his work during the March 5, 2007, teleconference. Neither of them recalled any such expression of support, and no expressions of support can be found in CASAC's March 26, 2007, letter review of the final Staff Paper (Henderson 2007b). The only relevant statement in this letter is a comment from panel member Sverre Vedal objecting to the statistical methods in the draft final EPA Staff Paper:

> [EPA's] approach amounts to attempting to find effects in a very few individuals when the statistical tests are not significant, which is a dangerous precedent – especially in this case where we are looking at small effects in 3 of 30 vs. 1 of 30, a pitiful number on which to attempt to base policy... (Henderson (2007b, p. C-30)).

The EPA staff is undeterred, however. Brown wraps his work in an imaginary CASAC endorsement. EPA staff then recycle Brown's unsupportable claim in the staff's Response to Comments (U.S. Environmental Protection Agency 2008e, p. 21) and, in the "voice" of Administrator Johnson, in the preamble to the final rule (U.S. Environmental Protection Agency 2008b, p. 16455/16451). Neither of these official Agency documents provides evidence that CASAC actually reviewed the matter beyond hearing a pair of three-minute presentations during its March 2007 conference call. Now that EPA has been challenged via our RFC, EPA has a very strong incentive to publicize such evidence if it exists, but in its Response to Comments EPA does not do so. EPA's response to our claim of information quality error is to attempt to cover it up by committing new information quality error.

3. Gerrymandering the scientific record

In response to EPA's reply, we've noticed that other public commenters expressed similar concerns about the possibility of systemic bias in the inclusion and exclusion of scientific studies. In a comment prepared on behalf of the Utility Air Regulatory Group (UARG), scientists at the Gradient Corporation identified 30 epidemiological studies published between 2000 and 2007 that EPA did not include in its scientific database (Gradient Corporation 2007, pp. A-1 to A-21). Consistent with the Iron Law, we have been unable to locate a single study that

---

study was prepared in response to the issues and analysis raised by a public commenter who made a presentation to the CASAC Panel at its March 5, 2007 teleconference."

arguably pushes the ozone risk envelope outward and was excluded from the scientific record.

There is ample evidence from both EPA documents and CASAC reports that both EPA staff and CASAC were primarily interested in research papers purporting to show a positive association between ozone and health effects. Thus, in addition to the problem of the "market supply" problem of publication bias (covered in Section IIIA.2 above), EPA's scientific record is contaminated by a matching "market demand" problem: only scientific evidence supporting EPA staff and predominant CASAC members' opinions about what policy the Administrator ought to choose were relevant to EPA's ozone review.

(a) EPA staff risk assessment methods show a preference for research showing positive effects

In December 2005, separate from the ozone review, EPA Deputy Administrator Marcus Peacock ordered a "top-down review" of the NAAQS standard-setting process. It appears that senior EPA officials had concluded that the existing process was not serving their needs. In the language of information quality, the process lacked adequate utility. Peacock's Memorandum does not reference EPA's Information Quality Guidelines, but nonetheless it refers to important information quality principles. For example, the Memorandum established as a presumptive norm that the EPA staff scientific record must be unbiased:

> The current NAAQS process has been in place for over 20 years, with some aspects required by law, and therefore not amenable to changes except through new legislation. Other important aspects of the NAAQS process, however, are discretionary -- the agency has established practices that set parameters for how science supports decision making. The Administrator is interested in determining whether those practices reflect the most rigorous, up-to-date, and unbiased scientific standards and methods (Peacock 2005, p. 1, emphasis added).

The Memorandum also reinforces the Administrator's desire that science be distinguished from policy in risk assessment, and in doing so strongly implies that EPA's Offices of Research and Development (ORD) and Air and Radiation (OAR) had persistently failed to make such distinctions. The assistant administrators for ORD and OAR were directed to establish a senior-level staff working group to solve this problem:

> In addition, the working group should focus on the nexus between scientific analysis and standard setting, including the degree to which we are successful in separating the exposition of scientific information from

> the development of risk management strategies and policy judgments (Peacock 2005, p. 2, emphasis added).

If senior EPA officials had been satisfied with the objectivity of the scientific information they were getting from Agency staff, there would have been no need to describe the initiative in these terms. In addition, they would not have encountered the strident opposition of current and former CASAC members, some of whom saw in the initiative a diminution of their ability to indirectly make policy decisions through their ostensibly scientific review function (Henderson 2008b; Vu 2005a, 2005b).

(b) Some CASAC panel members prefer research showing positive effects

CASAC members have not been shy about sharing strong policy preferences for more stringent NAAQS standards, and these views were known when they were recruited to serve on the panel. One CASAC member publicly opined that the ozone NAAQS ought to be more stringent and that the Administrator's most recent decision revising the particulate matter standard was illegal.[86] Another CASAC panel member participated in a process that in 2000 recommended an Air Quality Guideline for Europe of 120 $\mu g/m^3$ (~0.06 ppm) averaged over 8 hours,[87] and which in 2005 recommended that the AQG be lowered to 100 $\mu g/m_3$ (~0.05 ppm).[88] While these policy preferences often are

---

[86] Pinkerton et al. (2007): "To protect the nation's health, it is imperative that the EPA take action to issue a more stringent standard for ozone pollution." "We find the EPA posturing over scientific uncertainty to be disingenuous, uncompelling, and, ultimately, in violation of the Clean Air Act." The editorial acknowledges that co-author John Balmes was at the time a CASAC member who had been paid $52.80 per hour for approximately 25 hours of work over two years serving on the committee.

[87] See World Health Organization (2000, p. 33). A WHO Air Quality Guideline is similar to a primary NAAQS standard. It is a value that "provides a concentration below which no adverse effects or … nuisance or indirect health significance are expected, although it does not guarantee the absolute exclusion of effects at concentrations below the given value" (p. 42).

[88] See World Health Organization (2006, pp. 14-15). Note that the precision in these recommendations appears to be ± 10 $\mu g/m^3$ (~ 0.005 ppm). CASAC recommended that the Administrator set the primary NAAQS with precision ± 0.0005 ppm. See, e.g., the comments of Michael Kleinman on the 2nd draft Staff Paper (Henderson 2006c, p. D-33), and the CASAC letter opposing the Administrator's final decision (Henderson 2008a).

couched in scientific language, it is impossible to miss their policy content.[89] With regard to the scientific database, some CASAC members have openly called on EPA to include in the Criteria Document only those studies supporting the conclusion that ozone exposure below the current NAAQS poses significant human health risks[90] even though the panel as a whole advised EPA that "both positive and negative studies be given the same careful consideration."[91]

(c) Gray literature and "personal communications"

Several times in the Criteria Document, EPA cites as a scientific reference a conference or symposium presentation that was never published in a refereed journal.[92] It is especially noteworthy that EPA cites "Bell et al. (2006)" as the source for the strong claim that "if a population threshold existed for mortality, it would likely fall below a 24-h avg O_3 concentration of 15 ppb" (U.S. Environmental Protection Agency 2006a, p. 8-43). There is no scientific reference in the Criteria Document; "Bell et al. (2006)" is a personal communication between EPA staff and the lead author of an EPA-funded study who informally transmitted unpublished results in response to a staff query.[93]

---

[89] Kleinman's recommendation, cited in footnote 88, is clearly a mix of science and policy: "It would be appropriate to restate the current standard to 3 significant figures which is consistent with the precision of current monitoring devices and which will improve the margin of safety by eliminating 'rounding up' to 0.084." See also, e.g., comments by panel member Cowling, Lippmann, and Russell in Henderson (2005a). Cowling seems to have understood his job was to assist EPA staff in persuading the Administrator to endorse the staff's policy views; see Henderson (2005b, p. D-3). After the final rule was promulgated, CASAC sent what it called "unsolicited advice" stating that the Administrator's decision to set the NAAQS at 0.075 ppm was not "sufficiently protective" and characterizing their collective policy judgment as a "consensus scientific opinion" (Henderson 2008c, p. 2, emphasis added).

[90] See, e.g., individual comments by panel members Balmes, Lippman, and Miller, and the joint comment by panel members Legge, Hanson, Poirot, and Cowling in Henderson (2005a),

[91] See Henderson (2006a, p. 1; more detail on pp. 3 and 6).

[92] Gray literature in the Criteria Document includes Linn et al., 1983b; Lattimer et al., 1984; Selwin et al., 1985) Hogsett et al., 1989; Folinsbee and Hanucha, 1989; Spektor and Lippman, 1991; Tingey et al., 1991; Lebowitz et al., 1991; Linn et al., 1992; Laskin et al., 1996; and Sarwar et al., 2001.

[93] See (U.S. Environmental Protection Agency 2006a, pp. 7-179 and 178-183), citing "Bell, M. L. (2006) Community-specific maximum likelihood estimates of O_3-related excess risk in mortality for the NMMAPS U.S. 95 communities study [personal

In its Response to Comments, EPA claims that the Agency "can not include in its assessment results that were not reported" (U.S. Environmental Protection Agency 2008e, p. 33). EPA's reliance on unpublished data and results obtained through personal communications with Agency-funded researchers is inconsistent with that claim.

C. EPA Interprets and Presents Scientific Information in a Systematically Biased Manner

The Criteria Document, Staff Paper, and Notice of Proposed Rulemaking all collect, summarize and synthesize scientific evidence, much of it published in peer-reviewed journals. The challenge under applicable information quality guidelines is ensure that this information is accurate, reliable, and unbiased, and presented in an accurate, clear, complete, and unbiased manner. Each document displays evidence of both substantive and presentational bias, and bias appears to intensify in the progression from Criteria Document to Staff Paper to NPRM.

Interpretative bias arises in several forms. We discuss a few below.

1. *The inclusion or exclusion of data or studies based on the extent to which they support stated or unstated risk management objectives*

In our RFC, we said the inclusion of EPA's reanalysis of the Adams (2006a) data was evidence of purposeful bias because the reanalysis extracts selected data to jury rig support for Agency staff policy recommendations. It is a violation of information quality principles to choose a conclusion first, then fill in behind with selected data and contrived analysis to "support" it. A risk assessment performed this way cannot be unbiased, either in substance or in presentation.

In its Response to Comments, EPA "rejects" our contention that this is what Agency staff actually did (U.S. Environmental Protection Agency 2008e, pp. 21-22). We have already documented in Section III.B.2 the extraordinary efforts that EPA staff expended to rebut the statistical analysis in Adams (2006a). We also have shown that EPA's explanation for why it performed the reanalysis is false, that the Agency's recitation of the facts is both highly selective and self-serving, and that it has claimed CASAC's endorsement for its analysis despite CASAC review concluding before the reanalysis was completed. In short, EPA's

---

communication with attachments to Jee Young Kim]. New Haven, CT: Yale University School of Forestry and Environmental Studies; January 6." The Criteria Document also cites a personal communication with EPA-funded NYU assistant professor and EPA Science Advisory Board staff member Kazuhiko Ito (p. 7-185).

response to our claim of information quality error consists of disseminating new information quality errors.

2. *The inclusion or exclusion of data or studies based on post hoc or non-transparent criteria*

In our RFC, we objected to the EPA staff practice of drawing inferences from individual subjects in controlled human exposure studies when group mean effects are statistically nonsignificant (National Association of Manufacturers 2007, p. 19). The EPA staff's stated justification in the Staff Paper for cherry-picking Adams' data was a fishing expedition: EPA said "responses during the 0.06 ppm O_3 exposures appear to diverge from responses for filtered-air and 0.04 ppm O_3" in a manner that "is suggestive of a significant effect on FEV_1." EPA staff inferred that high interindividual variability combined with a "cursory evaluation" of Adams' newest data "strongly suggested that exposure to 0.06 ppm O_3 causes small group mean FEV_1 decrements in healthy adults with some individuals having notable effects," in this case FEV_1 decrements exceeding 10% (U.S. Environmental Protection Agency 2007f, pp. 3-8 to 3-9). EPA staff could not convert this "strong suggestion" into putative evidence until they discarded most of Adams data and applied their short-cut statistical procedure to the remnant.

In its Response to Comments, EPA repeats the fiction begun in the Criteria Document that its concern about the results in Adams (2006a) occurred because of apparently surprising interindividual variability in FEV_1 responses after 6.6 hours' exposure under exercise at 0.06 ppm (U.S. Environmental Protection Agency 2008e, p. 21). EPA has known about high interindividual variability in FEV_1 responses in controlled human studies for more than three decades. In the 1996 Criteria Document, EPA staff spent two pages summarizing interindividual variability observed in studies dating from 1972. In one EPA-funded study, ozone had accounted for only 31% of the of variance in FEV_1, "clearly demonstrating the importance of as yet undefined individual characteristics that determine responsiveness to O_3" (U.S. Environmental Protection Agency 1996a, p. 7-13). In the 2006 Criteria Document, no new research is cited attempting to explain this phenomenon; interindividual variability is simply characterized as "wide" and "considerable" (U.S. Environmental Protection Agency 2006a, pp. 8-16 to 18-18).[94]

[94] Intra-individual variation also appears to be large. Brown (2007a) analyzed a small subset of the data in Adams (2006a) in order to maximize the likelihood that differences in response could be interpreted as statistically significant. However, the

EPA categorizes individual FEV_1 decrements as "large" (20%), "moderate" (>10 but < 20%), "small" (3 to 10%), and "none" (± 3%) (U.S. Environmental Protection Agency 2006a, pp. 8-67 to 68-68). The group mean FEV_1 decrements reported by Adams (2006a) (and, incidentally, reproduced by EPA staff (Brown 2007a), fall within the "none" category irrespective of whether background is assumed to be 0.04 ppm (1.5%) or 0 ppm (2.8%). In a policy paper issued in 2000, the American Thoracic Society (ATS) noted the existence of a graded classification scheme for FEV_1 decrements issued by EPA in 1989, but commented that "[t]his classification has not been validated for acceptability or against other measures" (American Thoracic Society 2000, p. 671). ATS says nothing about EPA's 1996 graded scheme, which the Agency recycles in its 2006 Criteria Document.[95] Even if EPA's current graded scheme is assumed to be valid, only three of 60 subject-exposure pairs in the Adams' cohort experienced an FEV_1 decrement exceeding 10% after 6.6 hours of personal (not ambient) exposure to an average ozone concentration of 0.06 ppm -- one of 30 subjects for the square-wave test, and two of 30 subjects for the triangular-wave test (Brown (2007a, Attachment 1). These individual subjects – the only ones with responses in the "moderate" category -- drive the EPA staff reanalysis and provide the foundation for their reinterpretation of Adams (2006a) as displaying statistically significant decrements in FEV_1.[96]

3. Mischaracterization of results

In our RFC, we noted that scientific results can be misrepresented many ways, and we said several of these ways were evident in EPA's risk assessment documents.

---

correlation in responses after 6.6 hours exposure to 0.06 ppm ozone under the square- and triangular-wave protocols was only 0.48.

[95] EPA ignores the ATS caveat that EPA's 1989 scheme was not validated, then asserts that the Agency's latest graded scheme "appears to be valid and reasonable even in the context of the new ATS statement" (U.S. Environmental Protection Agency 2006a, p. 8-66).

[96] We are aware that CASAC encouraged EPA to cherry-pick data from the Adams cohort. The record shows that CASAC's motive was to advance its members' policy views. See Henderson (2006c, pp. 3-4, emphasis in original): "Adverse lung function effects were also observed in some individuals at 0.06 ppm (Adams, 2006[a]). These results indicate that the current ozone standard of 0.08 ppm is not sufficiently health-protective with an adequate margin of safety."

(a) Characterizing a study as "new" since the last ozone NAAQS review when in fact it was part of the last review

In our RFC, we said that EPA cited in this review many of the same studies the Agency had cited in its 1997 decision, but portrayed them as representing "new" scientific information.[97] For example, many of the controlled human studies EPA cites in the Criteria Document EPA also cited in the 1996 Criteria Document. These pre-1997 studies could be used now to say that the health risks posed by ozone are unchanged; after all, the scientific content of these studies cannot have changed. However, these studies cannot be used to support a claim that the health risks posed by ozone are more serious. Before using them to support a different risk characterization, EPA must show either that these studies contained previously unrecognized errors or that EPA had misinterpreted them. EPA did not do this; the EPA staff's "integrated synthesis" approach allows it subtly and nontransparently to reinterpret the scientific content of pre-1997 studies. Without such transparency, the public cannot test whether the EPA staff's portrayal of the science is substantively or presentationally objective.

We said in our RFC that it was misleading for EPA to confuse "old" and "new" scientific information in this manner. We said "EPA should segregate 'old' from 'new' science to ensure that the two categories are not confused, and discuss 'old' studies only to set the stage for its review of 'new' studies (National Association of Manufacturers 2007, p. 20). We noted that a reanalysis of an "old" study" constituted "new" science (p. 20, footnote 10). We also invited EPA to identify any pre-1997 study if "the Agency has learned about a material error" or discovered an error in its interpretation (p. 21).

In its Response to Comments, EPA says it "disagrees" with us "on both legal and scientific grounds" (U.S. Environmental Protection Agency 2008e, p. 157). EPA's legal ground for disagreement is facially suspect; the text consists solely of a restatement of the relevant law, including the very provision that most contradicts the Agency's position:

> Section 108 calls for the air quality criteria to 'accurately reflect the latest scientific knowledge useful in indicating the kinds and extent of all identifiable effects on public health or welfare' (U.S. Environmental Protection Agency 2008e, p. 157, emphasis added).

---

[97] For convenience, we use the term "pre-1997" as shorthand for those studies EPA included in the scientific database for the 1997 decision. We do not intend it to mean a literal dated demarcation.

Under EPA's interpretation of the law, the adjective "latest" is superfluous.

EPA's scientific ground for disagreement is that the staff chose an analytic framework that expressly permits it to reinterpret pre-1997 science differently from how they interpreted it in 1997:

> EPA implements this charge by reviewing the newest scientific information, and conducting this review not in isolation but by synthesizing and integrating the newest information with the prior scientific knowledge. An integrated synthesis of the entire body of evidence allows all of the evidence to be evaluated in context, without artificially segregating new from old information. It allows EPA to draw the most appropriate implications and conclusions from the evidence when seen as a whole (U.S. Environmental Protection Agency 2008e, p. 157).

This ignores the statutory context for decision-making under the Clean Air Act. The Administrator's task is to decide whether to revise the existing ozone standard, not to promulgate a brand new one. Thus, if the scientific record is going to have utility for that decision, it must segregate new from old information. Indeed, the rationale for the Administrator's proposed decision is segregated precisely this way: the Administrator first considered whether the existing standard was requisite to protect public health with an adequate margin of safety, and second the Administrator considered how much to lower it.[98]

EPA's Response to Comments further mischaracterizes the implications of distinguishing old from new science, and in doing so, makes surprisingly transparent the staff's desire to be able to reinterpret old science to meet new needs. Such a distinction would "call for freezing our understanding of the information gained from the 'old' studies," which is true only in part. Maintaining a clear distinction between "old" and "new" studies would in no way impede EPA staff from highlighting errors they have discovered errors in these 'old' studies, or identifying errors in their prior interpretation. What such a distinction would so, and which is highly desirable for substantive and presentational objectivity, is deter EPA staff from reinterpreting "old" studies in nontransparent ways.

Contrary to EPA's protests, such an approach is clearly "grounded in scientific principles" for it mimics almost exactly how scientists use and build

---

[98] In the preamble to the NPRM, the first step is set forth in Section II.C., with a conclusion in Section II.C.4. The second step is set forth in Section II.D., with a conclusion in Section II.E. See EPA (2007h).

upon prior literature. Never does the editor of a scholarly journal ask scientists to perform a de novo review of everything that precedes their submitted manuscript. Rather, they are required to summarize that literature briefly to provide a foundation for their work, and they are expected to clearly highlight any instance in which they believe the literature contains error or it has been incorrectly interpreted. EPA says our model is "neither required nor appropriate." Clearly, it is both.

(b) Characterizing a study as reporting something about which it is silent

In our RFC, we noted that in the NPRM EPA stated that results from "numerous" multi-city and single-city studies show that the associations between ozone and mortality "do not appear to be changed in multipollutant models including PM_{10} or $PM_{2.5}$ (U.S. Environmental Protection Agency 2007h, p. 37839). We noted that these "numerous" studies consist of the NMMAPS studies, and that the associations in these studies "do not appear to be changed" primarily because they do not measure $PM_{2.5}$.

We also noted two other examples of this form of bias: EPA's reanalysis of Adams (2006a), which we have already covered quite extensively, and EPA's misinterpretation of studies by Moolgavkar and coworkers (Moolgavkar 2000; Moolgavkar et al. 1995). We noted that Moolgavkar had disagreed with how EPA staff used his work (Moolgavkar 2007, pp. 4-5), and EPA staff has ignored these disagreements.

In its Response to Comments, EPA continues to deny that it has incorrectly interpreted Moolgavkar 's work (U.S. Environmental Protection Agency 2008e, pp. 54-55). EPA does acknowledge, however, having not included negative results reported by Moolgavkar – which is precisely the point.

(c) Characterizing a study as reporting something when it reports the opposite

In our RFC, we said EPA staff interpreted the literature as showing ambient ozone monitoring provided a satisfactory proxy for personal exposure. This is expressed most succinctly in the Staff Paper, which claimed that

> studies observed that the daily averaged personal O_3 exposures from the population were well correlated with ambient O_3 concentrations despite the substantial variability that existed among the personal measurements. Averaging likely removes the noise associated with other sources of variation. These studies provide supportive evidence that ambient O_3 concentrations from central monitors may serve as valid surrogate measures for mean personal exposures experienced by the population,

> which is of most relevance for time-series studies (U.S. Environmental Protection Agency 2007f, p. 3-41).

This is a strange construction. Whether average ambient ozone is correlated with average personal exposure matters only if the risks posed by ozone are the result of averages. Yet, the most peculiar aspect of the EPA staff claim is the authority cited to provide the scientific foundation. Neither of the studies referenced as the basis for this conclusion (Sarnat et al. 2005; Sarnat et al. 2001) actually make any such claim. Both studies say that ambient $PM_{2.5}$ but not ambient ozone is correlated with personal ozone exposure, and the researchers believe this is true because ambient ozone is a surrogate for personal $PM_{2.5}$. EPA asserts, but never explains, how these studies show that ambient ozone concentrations may serve as valid surrogates for personal ozone exposure.

(d) Selective and misleading citation

In our RFC, we provided numerous examples in which EPA cited CASAC selectively in the NPRM such that the result was a biased presentation of the panel's scientific review. We listed examples from the NPRM and added the text from the relevant CASAC document that EPA left out (National Association of Manufacturers 2007, pp. 22-27).[99]

In its Response to Comments, EPA "strongly denies" our claims and says that each of these issues was "thoroughly discussed in the NPRM" (U.S. Environmental Protection Agency 2008e, p. 152). However, each of the selective citations we listed came from the NPRM, making EPA's rebuttal technically infeasible. We agree wholeheartedly with EPA that it is "not required to quote verbatim all of an important comment made by the CASAC O_3 Panel," and that "[d]oing so in the Staff Paper or NPRM could have the effect of obstructing clear communication of the concepts involved rather than facilitating communication." However, the issue we raised was the lack of presentational objectivity in EPA's NPRM. Our complaint, which EPA does not rebut, is that the NPRM provides the public a severely biased and self-serving picture because it

[99] The task of discriminating between CASAC's scientific review and its policy advice is admittedly challenging. As we note in Section V beginning on page 131, this task was made immeasurably more difficult by EPA's decision not to ask CASAC to clearly distinguish between its scientific review and its policy advice, and CASAC's own decision not to be transparent about such distinctions.

quotes only the underlined text from a CASAC comment and excludes the context:[100]

- Since it is unlikely that each of these pollutants will have similar short-term effects on mortality, these findings suggest that while the time-series study design is a powerful tool to detect very small effects that could not be detected using other designs, it is also a blunt tool. The Clean Air Act requires that NAAQS be set for individual criteria air pollutants using the best available science. Because results of time-series studies implicate all of the criteria pollutants, findings of mortality time-series studies do not seem to allow us to confidently attribute observed effects specifically to individual pollutants. This raises concern about the utility of these types of studies in the current NAAQS-setting process and could serve to motivate interest in taking a broader perspective on regulating air pollution that incorporates the entire mixture of community air pollutants (Clean Air Scientific Advisory Committee 2006b, 3).
  - **EPA's defense is that it "addressed" CASAC's concerns in the Staff Paper and in Section II.D.4.a of the NPRM (U.S. Environmental Protection Agency 2008e, pp. 152-153). In the Agency's lexicon, to "address" a concern means to "discuss" or "consider" it, not to "resolve" or "reconcile" it.**
  - **The Staff Paper is irrelevant to our information quality complaint about the presentational objectivity of the NPRM, and the subsection of the NPRM EPA says is responsive appears 39 to 41 pages of dense Federal Register text later. A review of that text shows that EPA did not in fact "address" CASAC's concerns.**
  - **Whereas CASAC said the mortality studies "do not seem to allow us to confidently attribute observed [mortality]" to ozone, that is exactly what EPA did: "A standard set at [0.074 ppm] is estimated to reduce nonaccidental mortality [from ozone exposure] by**

[100] EPA also shifts the burden back to CASAC, with the added twist of insisting that it obey a warp in the space-time continuum: "If these issues had not been fully addressed, the CASAC O_3 Panel would have noted that in its final review of the Staff Paper, but it did not" (U.S. Environmental Protection Agency 2008e, p. 152). In short, CASAC's failure to propagate its every unresolved concern throughout each report in the series lets EPA off the hook. CASAC further failed to anticipate how EPA would cite its comments selectively in the preamble to the NPRM, which of course was published after CASAC's review was completed.

about 10 to 40 percent"(U.S. Environmental Protection Agency 2007h, p. 37877).

- Time-series studies typically make use of data from available air pollution monitoring network sites in which concentrations of various subsets of the criteria pollutants are measured. Study findings focus on identification of associations between day-to-day variation in these concentrations and daily mortality. Not only is the interpretation of these associations complicated by the fact that the day-to-day variation in concentrations of these pollutants is, to a varying degree, determined largely by meteorology, the pollutants are often part of a large and highly-correlated mix of pollutants, only a very few of which are measured. For the ozone and other photochemical oxidant NAAQS, this pollutant mix includes a large number of both gas- and particle-phase photochemical oxidant pollutants. Unfortunately, we have only limited information on the specific chemical composition, toxicity and, equally importantly, the population exposure of oxidant pollutants other than ozone (Clean Air Scientific Advisory Committee 2006a, p. 3).
    - **EPA's defense is that it "addressed" CASAC's concerns in several sections of the Staff Paper and in Section II.D.1 of the NPRM (U.S. Environmental Protection Agency 2008e, pp. 153-154). As indicated above, in EPA-speak "address" means "discuss" or "consider," not "resolve" or "reconcile."**
    - **The NPRM section referenced by EPA (U.S. Environmental Protection Agency 2007h, p. 37872), which appears 36 dense Federal register pages later than the selective citation from CASAC, is irrelevant. It consists of the Administrator's policy decision to retain ozone as the indicator for photochemical oxidant air pollution; it has nothing to do with CASAC's expressed scientific concerns about EPA's inferences about ozone-induced mortality.**

In our RFC, we said EPA had excluded from the NPRM crucial scientific comments from CASAC that did not support EPA's exposition of the data (National Association of Manufacturers 2007, pp. 25-27). We identified four such examples:

- The lack of correlation between ambient ozone levels (upon which all estimates of health risk depend) and personal exposures (upon which actual health risk must depend), especially among the elderly and infirm in which the alleged mortality effects from ozone are assumed concentrated.

- The inability to detect a threshold in concentration-response because of measurement error implied by the use of ambient ozone levels instead of personal exposures.
- The need for sensitivity analysis in the estimate of effects at different values for background, rather than the imposition of a policy-charged PRB.
- The possibility that ambient ozone serves as a surrogate for other pollutants, most notably $PM_{2.5}$.

For each issue we identified both the critical element of the CASAC comment and (unlike EPA) provided its context.

In its Response to Comments, EPA "rejects" our examples, saying that in each case the Agency 's exposition in the NPRM is complete and unbiased (U.S. Environmental Protection Agency 2008e, pp. 155-156). In each case, however, EPA actually cites irrelevant text from the Staff Paper, as if presentational objectivity in the NPRM is achieved as long as the Agency can point to text somewhere else in a subordinate document. EPA also cites pages in the NPRM where it says this material "can be found" or is "highlighted" (e.g., 72 *Federal Register* 37878), but EPA's reference concerns the Administrator's policy determinations, not an exposition of science.

(e) Drawing inferences from a study that are not supported by the data and analysis reported

In our RFC, we noted that EPA claims controlled human exposure studies provide compelling evidence that ozone exposure below the current ozone NAAQS causes lung function decrements, inflammation, and respiratory infection (National Association of Manufacturers 2007, p. 27). We also noted that the vast majority of the studies that EPA cites involve exposures at or above the current standard. EPA provided only a quasi-policy rationale for its ostensibly scientific inference, but that is impermissible under information quality principles. Policy officials have discretion over policy statements, but scientific statements must be supported by science.

In its Response to Comments, EPA denies that it drew inferences unsupported by the Adams (2006a) data (U.S. Environmental Protection Agency 2008e, p. 22). However, EPA does not deny that the group mean decrements in FEV_1 of 1.5% (compared to 0.04 ppm) and 2.8% (compared to 0 ppm) that Adams observed after 6.6 hours of exposure to 0.06 ppm is elsewhere characterized by the Agency as "within normal range (+3%)" (U.S. Environmental Protection Agency 2006a, p. 8-76, Table 8-2). EPA also does not deny that the sample standard deviation in FEV_1 responses after 6.6 hours' exposure to filtered air (i.e.,

zero ppm ozone) also was 3%.[101] EPA relies exclusively and completely on its eleventh-hour, never peer reviewed reanalysis of a selected fraction of Adams' data, having first tortured it to reveal statistical significance by discarding the experimental protocol in which the data were collected (Brown 2007a). EPA staff then characterize zero ozone exposure as "background" in order to try to nudge the FEV_1 decrement into its "small" effect size category (> 3%) – a threshold it still could not achieve without rounding 2.8% upward above the nearest integer.

Despite the obvious relevance and criticality of the EPA staff reanalysis of Adams (2006a) to the Administrator's policy determinations, the entire discussion of the reanalysis in the NPRM consists of a portion of a single paragraph, found at 72 *Federal Register* 37828, column 2). This discussion is peculiarly supplemented by footnotes (numbered 14-16) that are highly revealing. First, EPA uses passive voice to say that these results "were not included" in Adams (2006a). Second, EPA's reanalysis was truly an eleventh-hour work product (the memorandum for the docket is dated June 14, 2007, just six days before the Administrator signed the NPRM, and actually placed in the docket the same day as the NPRM). Third, what attracted EPA staff attention was that "7 percent" of Adams' subjects (i.e., two out of 30) experienced FEV_1 reductions after 6.6 hours exposure to 0.06 ppm ozone that they describe as "notable" (an undefined term), and then only when 0 ppm is used the presumptive.

Even this limited degree of transparency EPA provided grudgingly. The interagency review draft of the NPRM, dated May 22, 2007, contains no reference to EPA's reanalysis, which EPA had kept hidden since at least December 2006 when Brown alluded to results at a public meeting sponsored by EPA (Brown 2006). EPA even misleadingly tried to claim that its reanalysis was motivated by the need to "confirm" a public comment submitted to CASAC in early March 2007 (Smith 2007b). Smith tried had tried in vain to persuade CASAC to investigate more carefully the fundamentally flawed statistical analysis summarized in the Staff Paper.[102]

In our RFC, we said that EPA's analysis of clinical data on cardiac effects

---

[101] These data can be found in Brown (2007a, Attachment 1).

[102] The flicker of candor found in the NPRM appears to have been the product of interagency review. The story of EPA's secret reanalysis, and its unrelenting effort to mislead the Administrator and the public about its origin, is documented in Section III.B.2 beginning on page 63. The CASAC teleconference call at which Smith made his appeal occurred just three weeks before the court-ordered deadline for publication of the Staff Paper, so it may well be the case that by this time CASAC was helpless to act.

was similarly problematic with respect to information quality standards. The published studies show no statistically significant increases in dozens of endpoints examined, with one exception. In a study of 10 nonmedicated[103] hypertensive patients and six healthy adult males, approximately two dozen cardiac measures were obtained (Gong et al. 1998). Only two statistically significant differences were observed: a clinically nonsignificant 6% reduction in FEV_1 and a greater than 10 mm Hg increase in alveolar-to-arterial PO2 gradient ($AaPO_2$). In the NPRM, EPA emphasized the increase in AaPO2 and interpreted this as evidence that ozone exposure "result[s] in an overall increase in myocardial work and impairment in pulmonary gas exchange" (U.S. Environmental Protection Agency 2007h, p. 38734). EPA also was silent about the relevance of the exposure level (0.3 ppm, or 3.75 times greater than the current 8-hour NAAQS), or the uncertainties implied by extrapolating to the population clinical data obtained from a sample of 16.

In its Response to Comments, EPA does not "reject" our concern as it does so many times elsewhere (U.S. Environmental Protection Agency 2008e, p. 23). Instead, EPA says we should be mollified by other language it also used in which the Agency describes the cardiac data as "a very limited body of evidence" with "evidence for some potential plausible mechanisms." Re-reading the NPRM, we see that EPA characterized the cardiac epidemiology as providing "limited evidence suggestive of a potential association," which seems to us to be so qualified by caveats as to be meaningless if taken literally. The problem is that in the Administrator's statement of conclusions on the elements of the primary standard, these caveats are almost completely abandoned and "possibl[e] cardiovascular effects" are cited as evidence (U.S. Environmental Protection Agency 2007h, p. 37870). If scientific evidence this weak is considered "supportive" of a lower primary standard, it is difficult to imagine how weak evidence must be before EPA declines not to rely on it.

(f) Utilizing for one purpose data that were collected for another purpose

In our RFC, we objected to EPA staff's use of the Adams (2006a) data for purposes different that those which were intended by the study design (National Association of Manufacturers 2007, pp. 29-30). EPA staff first focused on the two of Adams' 30 subjects who had with the largest FEV_1 decrements after 6.6 hours of exposure to 0.06 ppm ozone under moderate exercise, and extrapolated their

[103] Although the abstract says the hypertensives were "nonmedicated," the text of the study describes them as "treated either pharmacologically for > 1 yr or by nonpharmacologic methods."

responses to the population. EPA staff then asked Adams for a highly restrictive subset of his data, and they proceeded to analyze these data without regard for Adams' study design. EPA staff never sought independent expert review of their analytic procedures, nor did they ever ask CASAC to review their work.[104]

In its Response to Comments, EPA defends this practice several ways (U.S. Environmental Protection Agency 2008e, pp. 97-98). First, EPA says that it performed a similar statistical analysis in support of its 1997 revised ozone standard. However, we have examined both the Criteria Document and the Staff Paper for the 1997 standard (U.S. Environmental Protection Agency 1996a, 1997), and we have found no evidence of an analogous statistical analysis, much less one utilizing paired *t* tests without adjustment for multiple comparisons.

Second, EPA says that it included data from Adams (2006a) because it was "urged" to do so by the American Petroleum Institute (API), which sponsored Adams' study. As we recounted earlier (see Section IIIB.2 beginning on page 63), the record shows that CASAC was first to ask EPA to include Adams (2006a); EPA neglects to mention this vital fact.[105]

Third, and most misleadingly, EPA asserts:

> The health risk assessment for lung function responses was reviewed by the CASAC O_3 Panel and there were no objections expressed by CASAC panel members or by Dr. Adams in either his oral or written comments to EPA concerning EPA's use of the Adams data as part of the basis for estimating the exposure-response relationships used in the health risk assessment (U.S. Environmental Protection Agency 2008e, p. 98, emphasis added).

---

[104] Only a summary is presented in the draft final Staff Paper.

[105] EPA's in-text reference in the Response to Comments for API's "urging" is "(API, 2006)," a reference not included in the bibliography. We infer that this reference is API's public comment dated September 18, 2006, on the second draft Staff Paper (Docket No. EPA-HQ-OAR-2005-0172-0057.1). This public comment says EPA "reasonably incorporates data" from Adams' studies, but notes with obvious concern that "rather than relying on these group mean results, the draft Staff Paper chooses to rely on data from individual subjects," a practice API correctly describes as statistically "invalid" (American Petroleum Institute 2006, pp. 19-20). EPA incorrectly states that Agency staff "obtained the individual data used in the health risk assessment directly from the author" [i.e., Adams] when in fact they sought only a very limited subset of the data set sufficient to perform its constrained statistical test.

Read carefully, is sentence refers only to the unobjectionable part of what EPA staff actually did. It should go without saying that the purpose of performing the research that became Adams (2006a) was precisely to help EPA "estimat[e] the exposure-response relationships used in the health risk assessment." That is a fundamentally different purpose, however, than the purpose for which EPA staff ultimately used it. The purposes of Adams' research were to (1) determine whether there were group mean decrements in pulmonary function at 0.06 and 0.04 ppm ozone, and if so, (2) determine whether these decrements differed by wave pattern. Adams (2006a) shows that pulmonary function decrements were not statistically significant at these lower concentrations, and that there was no difference in effect by wave pattern.

Had EPA staff allowed Adams' research to speak for itself, that would have been the end of the story. They didn't, and it wasn't. EPA staff cherry-picked from Adams' dataset, applied inappropriate statistical methods to make the selected data appear to show a statistically significant effect, and interpreted these results as compelling evidence of ozone health risk at 0.06 ppm. Adams publicly objected to this, and EPA attempts to cover up that fact.

(g) Hypothesizing after the results are known

In our RFC, we suggested that one of the information quality defects in the EPA staff approach is that it was hypothesizing after the results were known – a practice sometimes called "data mining" (National Association of Manufacturers 2007, p. 30). We are unable to locate any reply from EPA in its Response to Comments. EPA staff do not ever examine a health effect and attempt to discern its likely causes to estimate the fraction, if any, attributable to ozone exposure. No other factors matter, for ozone is the sole culprit of interest.

Properly performed hypothesis-testing requires researchers to specify *a priori* the hypotheses to be specified and the methods that will be used to test them. Improvisational data collection or statistical analysis after-the-fact are fine, but such research is properly described as either exploratory or hypothesis-generating, but never hypothesis-testing. The results of hypothesis-generating research should only be used to guide future hypothesis-testing research, and it never should be used to draw inferences – especially inferences that have significant public policy implications.

4. Study selection bias

In our RFC, we said EPA staff had displayed a systematic preference for studies that show positive associations even among studies that have important information quality limitations (National Association of Manufacturers 2007, p. 30). For example, where several studies were available to estimate effects on

asthmatics, EPA staff consistently selected studies with positive associations with ozone (e.g., Gent et al. 2003; Mortimer et al. 2002) over studies that do not (e.g., Moolgavkar 2000; Schildcrout et al. 2006). Nowhere in EPA's review plan, or in any other regulatory development document, did EPA discuss – much less establish -- an information quality basis for its selections. This bias is transparent when the EPA staff view of data from personal expiratory flow monitors is compared in the case of ozone ("data are reliable") and the case of nitrogen oxides ("data are unreliable"). The same studies are implicated; the only difference is that positive associations were obtained for ozone but not for nitrogen oxides.[106]

In its Response to Comments, EPA "rejects" our complaint, once again confusing having "discussed" or "considered" negative results and studies as equivalent to having taken them seriously (U.S. Environmental Protection Agency 2008e, pp. 55-56). EPA uses a "weight of evidence" approach that enables it to evade clarity and reproducibility – both hallmarks of good information quality practice. As we have noted, however, information quality principles and practices were missing from the ozone NAAQS review from beginning to end.

EPA also invokes as an all-purpose defense the fact that CASAC reviewed several of the documents subject to our RFC (U.S. Environmental Protection Agency 2008e, pp. 81-82). It is true, as EPA says, that the Agency's Information Quality Guidelines say that "if data are subjected to formal, independent, external peer review the information may generally be presumed to be of acceptable objectivity." EPA's guidelines presume, of course, that an agency's "formal, independent, external peer review" actually subjects the document to an information quality review. Peer review that ignores the information quality principle of objectivity cannot possibly ensure objectivity except by chance. As we documented in our RFC, and we reiterate here, EPA did not include information quality principles, in any shape or form, in its charge to CASAC.

EPA replies saying its 2005 Review Plan (U.S. Environmental Protection Agency 2005d) and first draft Health Risk Assessment (U.S. Environmental Protection Agency 2005b)[107] provided the "criteria for selection of studies and

---

[106] See EPA (2007b, p. 3-16) and the discussion in Section III.A.2(d)(vi).

[107] The in-text citation is to "Abt Associates, 2006." Abt is EPA's contractor. We infer that EPA intended to cite the October 2005 first draft Health Risk Assessment, which Abt produced and EPA published as if it were EPA's own work product. The document has footers on each page ascribing authorship to Abt, and it does not include a disclaimer stating that it was distributed solely for per review. Under the terms of

concentration-response relationships" (U.S. Environmental Protection Agency 2008e, pp. 81-82). However, the Review Plan actually contains no criteria for study selection. It is an outline of the review process and a description of the subjects to be addressed and nowhere mentions how studies would be selected for inclusion or exclusion. The first draft Health Risk Assessment contains a section titled "Selection of epidemiological studies" (4.1.5, p. 4-9), which lists the following criteria for study inclusion:

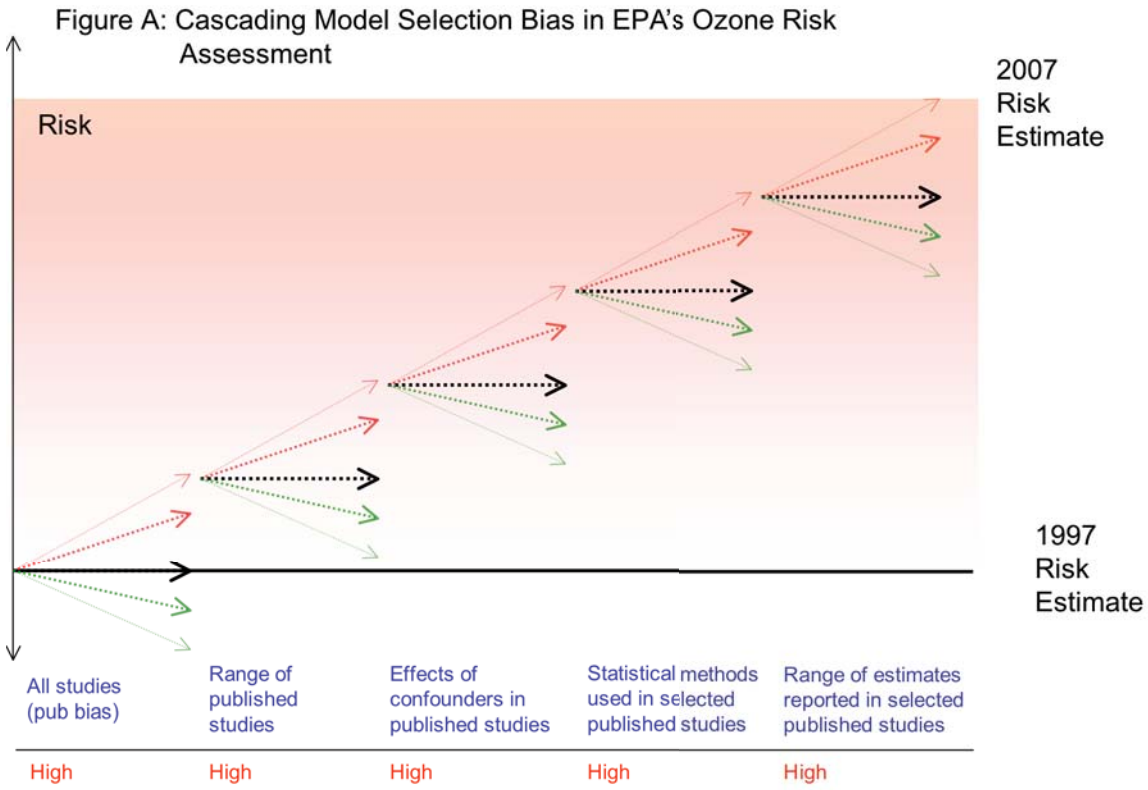
- It is a published, peer-reviewed study that has been evaluated in the draft O_3 AQCD and judged adequate by EPA staff for purposes of inclusion in this risk assessment based on that evaluation.
- It directly measured, rather than estimated, O_3 on a reasonable proportion of the days in the study.
- It either did not rely on Generalized Additive Models (GAMs) using the S-Plus software to estimate C-R functions or has appropriately re-estimated these functions using revised methods.
- For studies of mortality associated with short-term exposure to O_3, the study reported results for the O_3 season.

Information quality principles are missing from these criteria, and nonreproducible EPA staff judgment dominates.

5. Model selection bias

In our RFC, we said that EPA staff selected or emphasized models biased in favor of overestimating health risks:

- EPA staff selected models based on criteria other than quality of data or analysis
- EPA staff selected models known to yield upwardly biased risk estimates, such as single-pollutant models that do not control for known confounders
- EPA staff selected models based on statistically convenient but biologically implausible criteria
- EPA staff emphasized results from models known to yield risk estimates that are upwardly biased and more uncertain, such as



Generalized Additive Models conducted with insufficient convergence criteria

We used a figure, reproduced above, to illustrate the cascade of bias implied by just some of these practices.

(a) Selecting models based on criteria other than quality of data or analysis

We have already shown in the preceding section that EPA staff did not use information quality criteria to select models for risk assessment. In fact, the dominant criterion the staff used was its own non-transparent, non-reproducible, and undefined "judgment." In its Response to Comments, EPA staff defend their "judgment" noting that CASAC "did not express any concerns" about their choice of studies and models (U.S. Environmental Protection Agency 2008e, p. 82). This is hardly surprising, because the CASAC panel was dominated by researchers with strong policy views whose own work EPA was relying upon.[108]

[108] A notable example is Korrick et al. (1998), of which CASAC ozone panel member Frank Speizer is a co-author. Speizer cannot reasonably be expected to have

A peer review cannot be genuinely independent if it is conducted by the same scientists whose work the Agency is summarizing or promoting.

Several commenters recommended that EPA solve the model selection bias problem by adopting by Bayesian model averaging. EPA staff discussed this briefly in the Criteria Document, but they discarded it because it had certain undesirable effects – most notably, the magnitude of estimated effects would be "diluted (i.e., result in smaller coefficients) when variables are highly correlated, as may be the case for air pollution studies" (U.S. Environmental Protection Agency 2006a, p. 7-20).[109] EPA staff discuss the Bayesian model averaging study by Koop and Tole (2004) but discard it because its results "cannot be interpreted meaningfully." This test is related to information quality, but EPA staff do not subject the studies they rely upon to the same rigor. A more plausible explanation is that Koop and Tole found only very small effects. Meanwhile, the principle mortality study EPA staff rely on – Bell, McDermott et al. (2004) – uses a Bayesian data averaging procedure without which the authors could not have reported a statistically significant positive effect. Bayesian methods that push the ozone risk envelope outward are useful and appropriate; Bayesian methods that push the ozone risk envelope inward are not.

This is another practical example of EPA staff use of the Iron Law we presented in Section I.C. Scientific information indicating greater risk pushes the envelope outward; information that is equivocal supports the current location of the envelope; and information indicating lesser risk is discarded. We challenge EPA to refute the Iron Law by providing specific, concrete examples.

(b) Control for known confounders other than air pollution

In our RFC, we alluded to, but perhaps did not make clear, the fundamental information quality defect in the EPA staff's analytic approach. As an example, we listed 12 factors known to cause respiratory effects in asthmatic children: (1) air pollution, (2) cigarette smoke, (3), high humidity, (4) high/low environmental temperature, (5) allergens, (6) respiratory infection, (7), exercise, (8) nighttime hours, (9) stress or worry, (10) anger, (11) excitement, and (12) laughter (Sarafino et al. 2001). One way to manage this complexity is to try to estimate the contribution of air pollution (in this case, ozone) while controlling

---

given an objective review of this study or to have refrained from endorsing the EPA staff decision to give it a lot of weight. Moreover, it is difficult to imagine other members of the panel publicly identifying deficiencies in this study or opposing its use.

[109] EPA recycles these objections in its Response to Comments (U.S. Environmental Protection Agency 2008e, p. 44).

for these other factors, because failing to adequately control for known confounders yields upwardly biased estimates of risk. Moreover, such an approach would illuminate the characterization of adversity. It would be problematic, for example, to interpret as adverse physiological effects from ozone no greater than those from benign or positive phenomena. However, we noted that in the ozone epidemiology literature, control for confounders has been spotty, especially in ecologic studies but even in panel studies where individual data are obtained. It is remarkable, for example, that in the panel studies EPA relies on, there is no control for allergens and little control even for medication use.

EPA dismisses our concern about household allergens and exercise confounding the association between asthma and ozone exposure on the ground that they do not vary daily (U.S. Environmental Protection Agency 2008e, p. 42). We would agree if ozone's presumptive contribution to asthma were large relative to allergens, but it is not. Even according to EPA's preferred studies, however, ozone might be responsible for or exacerbate a tiny fraction of asthma cases, Most of the variance in asthma and its symptoms remains unexplained in these models.

Reflecting on EPA's Response to Comments, it now seems obvious that in practice the Envelope Theory requires that as much as possible of any health effect must be ascribed to air pollution (in this case, ozone). EPA staff do not seek to understand a specific health effect and try to discern the most plausible causes and allocate it objectively. The only time that EPA staff face a genuine dilemma is when multiple pollutants are effectively "competing" for a share of the air pollution burden, and in those cases EPA staff is susceptible to the temptation to assign the same health effects to each one.

This analytic defect carries over into the epidemiological research that EPA staff funds and, after publication, relies upon to advance its mission of tightening the NAAQS standards. We cannot find a single EPA-funded research study in EPA's scientific database that is focused on understanding the etiology of a health effect rather than seeking to prove that air pollution is its cause.

(c) Selecting models known to yield upwardly biased risk estimates, such as single-pollutant models and models that do not control for known confounders

In our RFC, we criticized EPA for basing its risk estimates on models known to yield upwardly biased estimates of health risk (National Association of Manufacturers 2007, p. 34). In its Response to Comments, EPA "rejects" our assertion that the Agency has done this, then proceeds to defend basing its risk estimates on models known to yield upwardly biased estimates of risk (U.S.

Environmental Protection Agency 2008e, pp. 44-45). EPA defends the use of single-pollutant models on the ground that they are "robust," or some similar formulation such as "fairly robust," "generally robust," or "statistically robust." EPA never defines any of these terms scientifically, but the Agency uses them in an ostensibly scientific context 54 times to describe associations in volume 1 of the Criteria Document, 48 times in the Staff Paper, and 28 times in the NPRM. We're not at all sure what EPA means by "robustness," but we do know that the Agency has not used the term in the same manner as it has been used by the statisticians who pioneered robust methods (e.g., Tukey (1977), Hoaglin et al. (1983).

Elsewhere in its Response to Comments, EPA mischaracterizes our complaint to imply that we think fully disclosing all relevant scientific information and results is a violation of applicable information quality, then proceeds to rebut its mischaracterization (U.S. Environmental Protection Agency 2008e, p. 83). We never made any such claim; we objected to EPA's reliance on models known to be upwardly biased for the purpose of human health risk assessment, not the comprehensive reporting of results. In short, EPA is violating information quality guidelines by purposefully estimating individual risk in a biased manner. With regard to our actual complaint, EPA is silent.

(d) In time series models, choosing lags based on statistically convenient but biologically implausible criteria

In our RFC, we objected to EPA's favorable treatment of several epidemiological studies in which researchers had mined the data to identify the most statistically significant lags and lag structures, then speculated why the results of these mining operations might be biologically meaningful (National Association of Manufacturers 2007, pp. 34-36). While we had no objection to the researchers' use of such exploratory data analysis techniques for EDA purposes, it was disconcerting to note that in some cases researchers drew inferences well beyond what EDA methods permit, and that EPA had treated these inferences as if they were confirmatory rather than exploratory.[110]

The time series studies EPA relies upon do not respect these fundamental biological requirements, and thus they sacrifice the weak presumption of objectivity they otherwise would enjoy under applicable information quality

[110] "Using techniques that adopt specifications on the basis of searches for high R^2 or high t values, is called data-mining, fishing, grubbing or number-crunching. This methodology is described eloquently by [Ronald] Coase: 'if you torture the data long enough, Nature will confess'" (Kennedy 1985, p. 76).

standards. Lags for specific health effects have been selected based on statistical strength without regard for the underlying biology, a procedure that yields upwardly-biased risk estimates (Moolgavkar 2007, pp. 6-7). Moreover, this has led to incoherence in lags across health effects, in which more severe health effects are implied to occur before milder ones.

In its Response to Comments, EPA "disagrees" with our characterization, even to the point of ignoring the actual statements of the researchers themselves, which make clear that their statistical analyses were exploratory in nature (U.S. Environmental Protection Agency 2008e, p. 45). EPA infers biological plausibility from statistical significance, rather than using statistical methods to test whether data are consistent with biologically plausible lags.

We emphasized EPA's reliance on Mortimer et al. (2002) as symptomatic of this constellation of information quality defects. The authors used seemingly every conceivable statistical device to discover positive associations: a wide array of lags and lag models; discarding statistically nonsignificant evening post-exposure effects in favor of statistically significant morning pre-exposure effects; then speculating about possible biological mechanisms that might explain their results. This is not controversial as an exercise in exploratory data analysis for the purpose of generating testable hypotheses, but it is completely inappropriate to interpret the results of EDA as confirming biological mechanisms concocted speculatively after the fact. In EPA's exposition, the exploratory nature of the researchers' data mining is downplayed and their results are treated as if they were confirmatory.

6. Assumption of causality

In our RFC, we faulted EPA for basing its conclusions about the causality of statistical associations on policy considerations rather than a plausibly objective scientific procedure. We illustrated a plausibly objective procedure (reproduced again as Figure B below) in which, *ceteris paribus*, effect sizes are treated the same regardless of their signs. We noted that EPA's approach consisted of putting a large policy thumb on the scientific scales:

> First, negative relative risk ratios are never suggestive of the absence of an effect. Second, positive relative risk ratios that are not statistically significant (and well below biological significance) are considered suggestive evidence of an effect. Statistically significant positive relative risk ratios are interpreted as suggestive evidence of a causal effect, and highly positive relative risk ratios are considered strong evidence of a causal effect.

> EPA's approach is generous with respect to interpreting positive associations as meaningful and quick to infer causality. This explains how EPA can collect many studies on ozone, each of which has small relative risks with small effects, and some of which are positive, and from this collection draw a "weight of evidence" conclusion that, when taken as a whole, the literature supports or strongly supports an inference of causality (National Association of Manufacturers 2007, pp. 38-39).

We added that EPA's approach is "unambiguously and transparently policy-directed" (p. 38).

In its Response to Comments, EPA "strongly disagrees" with the latter complaint, but provides only a boilerplate legalistic defense evading the point:

> The critical assessment of epidemiologic evidence presented in [section 7.1.2 of] the Criteria Document is conceptually based upon consideration of salient aspects of the evidence of associations so as to reach fundamental judgments as to the likely causal significance of the observed associations… (U.S. Environmental Protection Agency 2008e, p. 34).

This section of the Criteria Document is a discussion citing the Bradford Hill criteria and other "considerations" EPA staff took into account, including whether associations were "robust." As we noted in section III.C.5(c) on page 102, EPA uses the term "robust" and its variants a hundred times but never defines it.

Just as we said it was in our RFC, the EPA staff's process for determining causality is unambiguously and transparently policy-directed. EPA's Response to Comments appears to misinterpret our complaint to suggest that the Administrator or other policy officials directed the staff to embed policy judgments within their scientific review. We have not found any evidence suggesting such interference. Rather, we see a consistent pattern of EPA staff usurping the decision-making prerogatives of the Administrator and embedding their policy judgments into the science.