Attachment 1

John Engler

President and CEO

October 14, 2008

Ms. Molly A. O'Neill
Assistant Administrator, Office of Environmental Information
and Chief Information Officer
United States Environmental Protection Agency
Office of Environmental Information
1200 Pennsylvania Avenue, NW
Washington, DC 20460

Delivered via email to: mailto:quality.guidelines@epa.gov

Dear Ms. O'Neill:

On behalf of the National Association of Manufacturers (NAM), I am hereby submitting the attached Request for Reconsideration (RFR) in accordance with the procedures set forth in Section 8.6 of EPA's Information Quality Guidelines. The RFR concerns the Request for Correction (RFC) that the NAM submitted on October 9, 2007, logged in by your office as RFC #08001, to which EPA replied in its response to comments on the final revised ozone National Ambient Air Quality Standard (NAAQS). Per EPA's recommendation in Section 8.6 of its Information Quality Guidelines, I am attaching a copy of the RFC.

Among other issues, the RFR shows that many of the epidemiological studies EPA staff find persuasive used research designs that were known at the time to be demonstrably substandard. In some cases, EPA staff have relied on complex statistical methods to coax data into revealing effects from ozone so small that humans cannot even recognize experiencing them. Finally, EPA staff insist that certain studies provide valid and reliable evidence of respiratory health effects from ozone even though they rejected these same studies in their July 2007 draft Integrated Science Assessment for Oxides of Nitrogen -- and for the same reasons we mentioned in the RFC. Through the appeal, I seek more cogent answers than EPA provided in its response to the RFC. The document also identifies a number of process changes that are necessary to ensure that future NAAQS reviews fully and consistently adhere to the Agency's Information Quality Guidelines and the Information Quality Act.

The NAM appreciates the EPA's desire that stakeholders submit an RFR as promptly as possible and acknowledges the complexity of this issue area. The NAM has worked diligently to provide a document in a timely manner that articulates the association's concerns as thoroughly as possible. However, EPA's response to the RFC was scattered throughout both a 210-page Response to Comments document and the preamble of the final rule, which prolonged the analysis of EPA's response to the original petition, and therefore submission of the RFR.

Thank you for your consideration of the NAM's Request for Reconsideration. If you have any questions related to the attached RFR, please contact Bryan Brendle of the NAM staff at bbrendle@nam.org, or (202) 637-3176.

Sincerely,

John Engler

JE/blb

Attachments:
1) Request for Reconsideration
2) Request for Correction (filed with the EPA on October 9, 2007).

Table of Contents

I. Summary

A. Procedural Basis for this Request for Reconsideration

This Request for Reconsideration (RFR) is submitted in accordance with administrative procedures established by the U.S. Environmental Protection Agency (EPA) to ensure and maximize the quality of information the Agency disseminates:

> The Environmental Protection Agency (EPA) is committed to providing public access to environmental information. This commitment is integral to our mission to protect human health and the environment. One of our goals is that all parts of society - including communities, individuals, businesses, State and local governments, Tribal governments - have access to accurate information sufficient to effectively participate in managing human health and environmental risks. To fulfill this and other important goals, EPA must rely upon information of appropriate quality for each decision we make (U.S. Environmental Protection Agency 2002, pp. 47-49, emphasis added).

EPA is publicly committed to the principles of information quality. It is established Agency policy that:

- Disseminated information should adhere to a basic standard of quality, including objectivity, utility, and integrity.
- The principles of information quality should be integrated into each step of EPA's development of information, including creation, collection, maintenance, and dissemination.
- Administrative mechanisms for correction should be flexible, appropriate to the nature and timeliness of the disseminated information, and incorporated into EPA's information resources management and administrative practices.

On October 9, 2007, the National Association of Manufacturers, in adherence to procedures established by EPA in its 2002 Information Quality Guidelines, submitted a Request for Correction (RFC) contending that EPA's Notice of Proposed Rulemaking (NPRM) announcing the intent to revise the National Ambient Air Quality Standard (NAAQS) for ozone, and several supporting documents, each contained influential scientific information crucial for regulatory decision making under §§ 108 and 109 of the Clean Air Act that materially violated these standards (National Association of Manufacturers 2007). The RFC did not contest the statutory authority of the Administrator to make this decision; to make it promptly; or the nature of the criteria he was

required to take into account in making his decision. Indeed, the RFC makes clear that EPA's adherence to the information quality standards that the Agency committed to uphold was the best and surest way to fulfill this statutory mandate.

EPA responded to our RFC as part of its general response to significant public comments, which the Agency is required to prepare in compliance with Section 307(d)(6)(B) of the Clean Air Act. In its Information Quality Guidelines, EPA committed to integrate its responses to RFCs submitted in the context of regulatory actions within its regular responsibilities under the Administrative Procedure Act (and in this case, the Clean Air Act). We commend EPA for adhering to this important procedural element of its Information Quality Guidelines.

Unfortunately, EPA has not adhered to the substantive elements of its Information Quality Guidelines. Having carefully reviewed EPA's 210-page response, we can discern no instance in which the Agency conceded even the smallest of information quality error. Sixteen times, EPA said it "rejected" our concerns and complaints, often without any presentation of substantive data or argument. Twelve times EPA said it "disagrees" with us regarding the objectivity of a purported statement of fact, knowledge, or scientific inference, as if science can be reduced to a matter of opinion. Based on this review, we have concluded that it is necessary under the Information Quality Act to exercise our statutory right to seek and obtain the correction of error by means of the appeal procedures required by law and prescribed by EPA's Information Quality Guidelines.

EPA responded to our RFC two ways. First, EPA responded procedurally in a letter dated January 3, 2008, stating:

> The Administrator will issue his final decision on the ozone standards by March 12, 2008. At that time, EPA will respond to each of the issues raised in the RFC and other public comments received on the NPRM, either in the preamble to the final rule itself or in the accompanying Response to Comments document which will be placed in the rulemaking docket at the time the final rule is signed (Meyers 2008).

Second, EPA responded substantively to our RFC through the response to comments document that it normally publishes pursuant to Section 307(d)(6)(B) of the Clean Air Act (U.S. Environmental Protection Agency 2008d, hereinafter "Response to Comments") and in the preamble to the final rule. The preamble references NAM's public comment as an information quality RFC and mentions

it four times in the text.[1] The 210-page EPA Response to Comments cites the NAM public comment at least 45 times and twice notes that it was an RFC (pp. 150, 158).[2] Although we were the only public commenter to submit an RFC, EPA's Response to Comments includes responses to significant comments made by dozens of other public commenters.

EPA's decision, consistent with its Information Quality Guidelines, to incorporate its administrative procedures for managing information quality error correction requests within its normal rulemaking procedures, explains why EPA's document does not organizationally track our RFC. Indeed, sorting through the Response to Comments has been challenging. The document is redundant in places, and it includes comments ascribed to our RFC that we do not recognize having made.

This created some difficulty in determining which portions of EPA's Response to Comments are germane to our RFC. We have settled on what we think is a reasonable interpretative strategy:

- Where EPA's Response to Comments mentions a comment that it ascribes to public commenters other than NAM or to unidentified commenters, we interpret this to be not part of EPA's response to our RFC.
- Where EPA's Response to Comments mentions a comment that it ascribes to multiple commenters using the form (ABC, XYZ, NAM) without page numbers, we interpret this to not be part of EPA's response to our RFC unless the issue at hand is strictly scientific.[3]
- Where EPA's Response to Comments mentions a comment by NAM using the format (NAM, p. *x*), we interpret this to be a formal response to the RFC.

---

[1] EPA (2008b, pp. 16454, 16457, 16466, and 16469).

[2] A search of the EPA docket reveals a 72-page document styled as a response to the NAM RFC. However, this document also includes responses to many other public commenters. We infer that this other document is not authoritative but was placed in the docket only because it was shared with the Office of Management and Budget. See (U.S. Environmental Protection Agency 2008c).

[3] Separately from the RFC, which is strictly limited to scientific, statistical and technical matters covered by applicable information quality guidelines, NAM also submitted a traditional public comment that addressed a broad array of issues including policy considerations that are not subject to applicable information quality guidelines.

This RFR is limited to matters within this third category.

Finally, although the text of EPA's January 2008 letter advised us that EPA might respond to our RFC within "the preamble to the final rule itself," we cannot find any text in the preamble that reasonably can be construed as a response to the RFC as opposed to a response to comments more generally.[4] Therefore, in this RFR we focus on EPA's Response to Comments as the authoritative EPA response to our RFC. For everyone's convenience, we follow the structure and organization of our RFC (which contains only information quality-related issues) rather than EPA's Response to Comments (which contains responses to all significant comments, including comments made by others and a large number of comments on policy).

B. EPA's Response to Comments Offers No Evidence that the Agency Adhered to Its Own Information Quality Principles, Policies and Procedures

EPA's Response to Comments proves beyond any reasonable doubt that until we submitted our RFC, EPA staff, management, and policy officials had devoted no attention to information quality in the revision of the ozone NAAQS. In every EPA staff document, beginning with the Review Plan (U.S. Environmental Protection Agency 2005e), proceeding to the Criteria Document (U.S. Environmental Protection Agency 2006a, 2006b, 2006c), the Exposure Assessment and Risk Assessment (U.S. Environmental Protection Agency 2005a, 2005c, 2006g, 2006h, 2006i), and the Staff Paper (U.S. Environmental Protection Agency 2006f, 2007j), there is no mention, discussion, analysis or any other content mentioning, discussing or applying the requirements of the Information Quality Act, the government-wide implementing guidance issued to all agencies by the Office of Management and Budget (Office of Management and Budget 2002), or EPA's own implementing guidelines (U.S. Environmental Protection Agency 2002, 2003, 2006e). EPA's Information Quality Guidelines require that Agency program offices perform sufficient pre-dissemination review to ensure that the quality of information that is disseminated is maximized. However, the Response to Comments shows that EPA staff performed no pre-dissemination review of the information quality aspects any of the scientific information that they transmitted to the Administrator in support of his policy decision.

---

[4] At 73 Fed. Reg. 16510, EPA cites our submission as both a "letter" (i.e., a "Request for Correction") to EPA Assistant Administrator (Environmental Information) Molly A. O'Neill and as "public comment" to Docket No. EPA-HQ-OAR-2005-0172. Some text in the preamble to the final rule is essentially identical to text in the Response to Comments.

By law, EPA's technical staff work products must be peer reviewed by the Clean Air Scientific Advisory Committee (CASAC). EPA peer review guidance, which specifically covers peer reviews such as the one performed by CASAC, commit the Agency to ensure that peer reviews fully address information quality issues (U.S. Environmental Protection Agency 2006e). However, information quality was omitted from the panel's charge. CASAC meetings are dialogues between panel members and EPA managers and staff, yet the transcripts of each in-person meeting shows that neither the principles nor the procedural and substantive requirements of information quality were ever mentioned by any EPA manager or staff member.

The absence of information quality from every aspect of the ozone NAAQS review is complete and comprehensive. Yet in its Response to Comments, EPA "rejects," "disagrees" with, or otherwise denies each and every information quality error claim in our RFC:

> EPA has reviewed NAM's RFC and finds that there is no merit to their objections. EPA disagrees with NAM's allegations that EPA has not complied with the requirements of the Information Quality Act or the Agency's policies for ensuring information quality. EPA has responded to NAM's significant comments in the preamble to the final rule or in this document (U.S. Environmental Protection Agency 2008e, p. 158).

EPA staff devoted thousands of man-hours and millions of dollars to assemble, summarize, analyze, and write thousands of pages of scientific reports for this review. We have found not a single page that concerns information quality.

This RFR responds to EPA's replies in the same fashion that the RFC was written. Where EPA has provided persuasive evidence that it is correct or that our evidentiary case is insufficient, we withdraw our request for correction. Where EPA's reply is problematic, however, we have summarized or restated our claims and put them forward again on appeal. In reviewing EPA's replies, we have discerned certain patterns. Many of EPA's replies fall into one or more of the following categories:

- EPA has mischaracterized our information quality claim, often in the form of a straw man, and responded to its mischaracterization rather than our claim.
- EPA has mischaracterized our information quality claim as a matter of opinion, as if representations of knowledge such as facts and data can be subjectively determined, then asserted that its opinion is superior to ours.

- EPA has mischaracterized a scientific issue as one determined by law or policy judgment.
- EPA has characterized our information quality claim accurately, but responded to an irrelevant or unrelated issue or merely responded with boilerplate.
- EPA has responded to our complaint of information quality error by committing a new information quality error, typically by making new informational statements that fail the substantive and/or presentational objectivity standards.

C. The Iron Law of EPA Staff Ozone Health Risk Assessment and Characterization

A comprehensive review of our information quality error correction claims and EPA's responses has led us to discern a Iron Law of EPA Staff Ozone Health Risk Assessment and Characterization that explains how EPA staff utilize scientific information. The Iron Law is set forth in the nearby text box, and we refer to it frequently in this RFR. The scope, scale and magnitude of risk can be understood as an envelope or a balloon; higher risk means a more expansive envelope or a larger balloon.

Science suggesting the potential for greater risk pushes the risk envelope outward or adds air to the balloon. Science that is equivocal supports the envelope at its current location or maintaining the balloon at its current size. Science suggesting lower risk moves the envelope inward or removes air from the balloon, but EPA staff will use such information only under conditions that are so restrictive as to be nearly impossible to meet. Science that does not meet these conditions is "discussed" or "considered," but ultimately discarded. The principles of information quality play a severely constrained role: they are used only as barriers to the admission of evidence indicating lower risk.

Ironically, the foundation for the Iron Law of EPA Staff Ozone Health Risk Assessment and Characterization was first set forth by EPA staff itself, in a 2004 report titled *An Examination of EPA Risk Assessment Principles and Practices* (U.S. Environmental Protection Agency Office of the Science Advisor 2004b). In that report, EPA staff elucidated publicly for the first time that it is the policy of EPA staff not to understate risk or to grossly overestimate it.

Text Box 1:
The Iron Law of EPA Staff
Ozone Health Risk Assessment and Characterization

1. EPA staff use new scientific information suggesting greater potential or actual health risk as evidence that risk is greater than previously believed. EPA staff make no practical distinction between health risk that is potential (i.e., possible, hypothetical, speculative) or actual (i.e., proved).
2. EPA staff use new scientific information that is equivocal to support their existing assessment and characterization of health risk.
3. EPA staff use scientific information suggesting lesser potential or actual health risk as evidence that risk is lower than previously believed, provided that its quality is flawless in every respect.
4. EPA staff use scientific information suggesting the absence of human health risk only if it proves that risk is biologically infeasible.

D. Types of Major Information Quality Errors in the Scientific Record for the Ozone NAAQS Review

This RFR documents a long list of information quality errors, but some errors clearly are more significant and important than others. We present a Baker's Dozen below:

1. *EPA omitted any reference to information quality principles and own Information Quality Guidelines from every document in the ozone NAAQS review, stretching from the 2005 Review Plan to the 2007 NPRM.*

EPA's Information Quality Guidelines commit the Agency to instill information quality principles and practices throughout its regulatory development process. Yet, information quality is completely missing from the ozone review. It is impossible for EPA to simultaneously have adhered to information quality principles yet have been utterly silent about them.

2. *EPA "considers" and "discusses" a phenomenal quantity of scientific information, but only uses information in accordance with the Four EPA Staff Principles.*

For many of the information quality issues raised in the RFC, EPA says in its Response to Comments that it "considered" or "discussed" it, usually in the

Criteria Document. While it is certainly important for EPA staff to have done these things, they are not the same thing as having examined and evaluated the information quality attributes of scientific information and accounted for information quality throughout the review. The comprehensiveness of EPA's discussion was not the focus of our RFC; we objected to the lack of objectivity in these documents.

We are unable to locate a single scientific study that both pushes the ozone risk envelope outward and was excluded by EPA. Similarly, we cannot identify a single scientific study that EPA puts any weight upon which pushes the ozone risk envelope inward.

3. *EPA makes crucial claims that are easily refutable.*

The quintessential example is EPA's untruthful explanation of the origin for its reanalysis of selected data from the controlled human study by Adams (2006a). EPA claims that its reanalysis was prompted by public comments to CASAC provided by Smith (2007b) in March 2007. We prove that EPA's reanalysis was substantially completed by December 2006. EPA declined to distribute this work for timely public comment, and instead hid it from peer review by CASAC and placed it in the docket the same day that the Administrator signed the proposed rule. EPA also falsely claims that CASAC supported its reanalysis despite the absence of any CASAC review.

EPA's reanalysis (Brown 2007a) was a crucial element of the scientific record on which the Administrator relied to decide what ambient concentration of ozone is requisite to protect public health. EPA's description of its reanalysis of the Adams' data, culminating in the eleventh-hour insertion of Brown (2007a) into the scientific record without CASAC peer review, is deeply defective with respect to presentational objectivity.

4. *EPA uses ad hoc statistical analyses devised after the data were obtained.*

EPA has relied on controlled human exposure studies since at least the 1997 ozone NAAQS review. Indeed, the Agency has made significant investments in facilities and staff to perform controlled human studies. These studies have always been carefully designed (if not flawlessly implemented), and until now EPA has followed recognized and accepted statistical procedures for analyzing data.

For the first time, however, a controlled human study did not reveal statistically significant decrements in pulmonary response. In response, EPA first discarded the author's portrayal of his results, focused on selected observations from individual study subjects, and reanalyzed a gerrymandered subset of the

data to "prove" that the decrements observed are statistically significant after all. Through the vehicle of Brown (2007a), EPA has dispensed with longstanding statistical practice. When challenged, EPA staff defend this by incorrectly claiming that their statistical procedures are more commonly used than are those of the researcher whose analysis they reject.

5. *EPA disseminates risk characterizations based on epidemiological studies that use unvalidated self-reported data collected in diaries.*

Several of the epidemiological studies EPA staff rely upon are based on data obtained from diaries kept by study subjects or their caregivers. It has been shown that this research design results in self-invented data. Such data are unreliable unless significant proactive steps are taken, both in the recruitment phase and during study implementation, to ensure reliability and accuracy. These steps were not taken by the researchers who authored the studies that EPA uses to base its characterization of respiratory risks (Gent et al. 2003; Mortimer et al. 2002; Ross et al. 2002).

6. *EPA disseminates risk characterizations based on pulmonary function data obtained through a low-resolution clinical diagnostic procedure that cannot reliably or accurately detect subtle effects.*

The pulmonary function tests that epidemiologists use were intended for the clinical purpose of diagnosing disease and assigning patients into rough categories. They were never intended for measuring or estimating very small changes within individuals or across populations. Clinicians are trained to coach patients in their performance, a procedure that is reasonable for medical evaluation but improper for research purposes. When investigators are not blind, or multiple investigators with even subtly different coaching techniques are involved, the opportunities for error and bias are legion.

Several of the panel studies that EPA uses to characterize respiratory risk rely on these unreliable methods (Gent et al. 2003; Korrick et al. 1998; Ross et al. 2002). That is the case here, where epidemiologists seek to detect low single-digit percentage differences in pulmonary function and attribute these differences to air pollution.

7. *EPA disseminates risk characterizations based on epidemiological studies of pulmonary function in which the research design discards variability and uncertainty, thus making association with air pollutants appear to be much more certain than they actually are.*

Some of the epidemiological studies EPA staff rely upon are based on pulmonary function tests that measure phenomena that are both uncertain and variable. Researchers obtained data for multiple test maneuvers conducted at the

same time, but discarded data judged to be "unacceptable" and recorded only the average of the values not discarded. This practice violates a crucial assumption underlying the statistical tests that the epidemiologists subsequently performed. Assuming that uncertain phenomena are fixed when they are not reduces estimated standard errors, exaggerates statistical significance, and misleads decision-makers and the public about uncertainty and precision.

In studies of asthmatics and others with compromised pulmonary function, inter-maneuver variability can be very large. Assuming that inter-maneuver variability is zero yields unreliable and artificially narrow standard errors, and inflates statistical significance. It is virtually certain that discarded inter-maneuver variability exceeds in magnitude the small percentage pulmonary function decrements observed in these studies.

The consequence of discarding inter-maneuver variability is least important in controlled human studies of homogeneous subjects at high ozone concentrations. In these studies, inter-maneuver variability is low due to strict selection criteria for study subjects, and relatively large effect sizes are expected. Nevertheless, as the exposure concentration approaches background and the expected effect size approaches zero, the importance of discarded variance increases and almost certainly swamps the effect size. In the case of the only controlled human study to test for pulmonary function under exercise through 6.6 hours of exposures to 0.04, 0.06, and 0.08 ppm (Adams 2006a), study subjects provided pulmonary function data from at least two maneuvers such that the sum of FEV_1 and FVC was ± 200 ml, or about 3% on average. This variation alone is more than half the group mean pulmonary function decrement observed at 0.08 ppm, and is more than twice the group mean pulmonary function decrement observed at 0.06 ppm. Once inter-maneuver variability is accounted for, the minimum effect size that is truly detectable in controlled human studies may well be larger than the difference between ambient and background ozone levels.

8. *EPA disseminates risk characterizations based on epidemiological studies of pulmonary function in which the research design requires the use of biased estimates.*

Some of the epidemiological studies EPA staff rely upon for risk characterization are based on pulmonary function tests in which the fixed value recorded is the largest value obtained from a series of maneuvers performed in close sequence, not the average (Korrick et al. 1998; Mortimer et al. 2002; Ross et al. 2002). As indicated above, in clinical practice subjects are routinely coached to perform so that they achieve maximum results. Maximum performance is especially sensitive to coaching effectiveness (better coaches produce higher

maxima) and the number of maneuvers performed (more maneuvers increase the expected maximum). EPA staff have not acknowledged, much less analyzed, the consequences of bias in the pulmonary function testing performed by epidemiologists.

9. *EPA disseminates risk characterizations based on epidemiological studies of unrepresentative samples, or samples whose representativeness has not been validated.*

Several of the epidemiological studies EPA staff rely upon for risk characterization use convenience samples (Korrick et al. 1998) or cohorts whose representativeness has not been shown (Gent et al. 2003; Mortimer et al. 2002; Romieu et al. 1996). Convenience samples are presumptively unrepresentative. The representativeness of study cohorts that were assembled by non-randomized designs cannot be presumed. EPA staff rely upon coefficient estimates obtained from these studies and simply assume that the underlying samples are representative.

10. *EPA disseminates risk characterizations based on epidemiological studies with unaccounted for or unreported nonresponse bias.*

For at least two decades, federal statistical policy has required that surveys and similar studies achieve response rates that epidemiologists typically find problematic, and this policy has recently been codified in formal government-wide guidance (Office of Management and Budget 2006). Any federally-sponsored study – a term that generally includes EPA-funded epidemiology -- that is not expected or fails to achieve an 80% response rate must include a thorough nonresponse bias analysis. In the epidemiological studies EPA staff rely upon for risk characterization that have individual data, none achieved an 80% response rate and nonresponse bias analyses were either not performed or performed but not disclosed. Publication in a refereed journal confers a rebuttable presumption of "adequate" objectivity, but this presumption is automatically rebutted in any case where response rates do not satisfy applicable federal statistical policy standards and a rigorous nonresponse bias analysis was not performed.

11. *EPA disseminates risk characterizations based on ambient monitoring data as a proxy for personal exposure despite very low correlation.*

Several of the epidemiological studies EPA staff rely upon use ambient ozone data obtained from central monitoring sites as proxies for personal exposure despite overwhelming evidence that ambient and personal exposures are poorly correlated (Bell, McDermott et al. 2004; Gent et al. 2003; Mortimer et al. 2002). Despite this lack of correlation, EPA staff interpret observed weak

positive associations between ambient ozone and various health effects as causal. The best that can be said scientifically about these studies is that ambient ozone monitors are measuring something that might be associated very weakly with morbidity, mortality, and other phenomena such as emergency department visits, hospital admissions, and school absences. EPA staff assume that ambient ozone concentrations are functionally equivalent to personal exposure and that the observed weak associations are causal, but they cannot reach either inference based on science.

EPA staff go even further to assert that the fact personal exposures are tenfold or more lower than ambient levels supports the assumption that the health risks posed by ozone are underestimated. There are two problems with this claim. First, it doesn't matter what their ratio is if they are uncorrelated. Second, it turns algebra on its head. In controlled human studies, subjects were subjected to personal – not ambient – concentrations of 0.06 ppm. EPA acknowledges that ambient ozone concentrations are 2- to 4-fold greater than personal exposure. That means the 1.5% group mean FEV_1 decrement observed by Adams (2002, 2006a) at 0.06 ppm personal exposure corresponds to an ambient ozone concentration of 0.12 to 0.28 ppm.

12. *EPA disseminates risk characterizations based on studies using research methods Agency staff have rejected as unreliable and invalid for other air pollutants.*

Several of the epidemiological studies EPA staff rely upon in the ozone NAAQS use personal expiratory flow rate devices. For ozone, EPA staff say these methods obtain valid and reliable data. However, in their draft Integrated Science Assessment for Oxides of Nitrogen (NO_x), EPA staff say these devices are known to yield unreliable and invalid data. In the ozone review, EPA staff highlight Mortimer et al. (2002) as especially relevant for the estimation of health risks to asthmatics. In the NO_x review, EPA staff highlight Mortimer et al. (2002) as an example of a study whose methods are irredeemably deficient.

The Iron Law of EPA Staff Ozone Health Risk Assessment and Characterization clearly are at work here. Mortimer et al. (2002) reported statistically significant positive associations between ambient ozone and reported symptoms, but no association between those same symptoms and NO_x. Hence, EPA staff use Mortimer et al. (2002) to push out the ozone risk envelope, and they discard it rather than use it to push the NO_x risk envelope inward.

13. EPA never informed CASAC about information quality principles or the Agency's Information Quality Guidelines, and information quality played no role in CASAC's review even though it is required by EPA's Peer Review Handbook.

By law, CASAC has the unenviable task of simultaneously providing both an objective review of the scientific database and policy advice to the Administrator that, by its very nature, cannot be objective. EPA could have made this task easier if it had asked CASAC to clearly distinguish its scientific review from its policy advice. Indeed, information quality review is an explicit component of EPA's Peer Review Handbook. It is essential for peer reviews to fully address information quality principles in order to secure the rebuttable presumption of objectivity that the Information Quality Guidelines provides.

However, EPA didn't inform CASAC about information quality, nor did the Agency's Charge to CASAC even mention the subject. Consequently, the record of the CASAC review shows that the panel wove its scientific review and policy advice into a single fabric and, predictably, paid no attention to information quality. It is infeasible for CASAC's review to have fully addressed information quality principles when it devoted no time at all to information quality issues.

E. Remedies Requested

The evidence for systemic information quality error by EPA staff is overwhelming, both procedurally and substantively. The difficulty EPA staff face is that a candid, accurate and forthright response to this RFR may undermine the Agency's ability to legally defend the Administrator's recent decision. We have tried throughout this process to stay clear of the Administrator's exercise of discretion, as provided for by the Clean Air Act, but we agree that EPA's ability to defend is highly compromised by the existence of systemic information quality error that rendered inaccurate the scientific database on which the Administrator relied. Moreover, the Clean Air Act imposes on EPA the onerous duty to revise each NAAQS every five years despite the fact that it takes about that long to conduct each review. That means EPA is engaged in a never-ending cycle that impedes it from implementing the process reforms necessary to comply with its own information quality policies and guidelines.

Senior EPA officials are obligated under Section 8.6 of the Agency's Information Quality Guidelines to perform an independent review of this RFR and provide a well-documented and comprehensive response to each information quality error that we continue to allege. In some cases, it may not be possible for EPA to repair the error we have identified in the current NAAQS

review cycle. In no case, however, should the inability to make a timely repair justify continued misrepresentation of the information in the scientific database.

In addition, we specifically request that EPA make significant changes to its NAAQS review process so that future information quality errors are rare instead of systematic.

1. *Explicitly, comprehensively, and consistently implement information quality principles and practices throughout the NAAQS review process and within each NAAQS work product.*

The record shows that EPA ignored information quality principles and the Agency's own policies and procedures throughout the ozone NAAQS review. Agency officials must completely overhaul the NAAQS review process to explicitly, comprehensively, and consistently implement information quality. Token efforts, changes made only at the fringe of the process, reforms that shift EPA's burden to the public, and the addition of new legalistic boilerplate are all unacceptable.

2. *Establish an external independent body with the limited responsibility of reviewing the quality of scientific and technical information, and advising the Administrator whether information quality principles have been met and applicable information quality policies and practices have been followed.*

To prevent information quality error, we recommend that the Administrator establish an external and independent Information Quality Review Committee for the express and limited purpose of advising whether information quality principles have been met and information quality policies and practices have been followed. The Committee would supplement, not supplant, the scientific and policy review currently performed by CASAC.

The Committee would not be asked to make policy recommendations or opine on what the science means, both of which are statutory functions currently assigned to CASAC, but to perform information quality reviews that are significantly different from CASAC's current activities. These functions cannot be performed by CASAC because its members generally lack expertise in information quality, and it may be unreasonable to expect them to have both information quality and subject matter expertise. Moreover, they are pressed for time to satisfactorily accomplish their current assignments.

The work of the Information Quality Review Committee should be performed in public subject to the requirements of the Federal Advisory Committee Act. Committee members must be independent of EPA and unaffiliated with the research teams whose scientific work products the Agency

relies on for risk assessment and characterization. It is impossible for any individual or panel of peer reviewers to independently examine the quality attributes of research that they personally performed or that was performed by their institutional colleagues.

3. *Remove all policy judgments and similar considerations from the assembly, review and presentation of scientific information in all EPA NAAQS work products.*

The record shows convincingly that EPA staff routinely attempt (and in the ozone NAAQS review, appear to have succeeded) in restricting the authority delegated to the Administrator by the Clean Air Act through the device of providing the Administrator with a summary of the scientific database that reflects EPA staff views about decisions they believe the Administrator ought to make. This can only be overcome if EPA officials explicitly and forcefully direct the staff to refrain from embedding policy judgments in these work products and instead provide the Administrator with a genuinely objective scientific record. Instructions to CASAC – the "Charge" – also must be modified so that the committee is explicitly and formally directed to clearly distinguish its scientific review from its provision of policy advice.

The Information Quality Review Committee should be tasked with determining whether this directive has been met, and if it hasn't, informing the Administrator where residual policy judgments reside. Relentless effort is needed to ensure that science and policy are clearly distinguished in all EPA NAAQS work products.

4. *Establish new and publicly accountable pre-dissemination review procedures for all EPA NAAQS work products.*

The record shows that despite the existence of an Agency requirement for pre-dissemination review that was established in 2002, no such review ever took place in the current ozone NAAQS review. It is not credible to believe that staff were unaware of information quality principles and Agency guidelines.

EPA officials should rectify this apparent loophole by explicitly establishing a comprehensive program of pre-dissemination review of the information used in NAAQS reviews. These activities can be conducted in parallel with NAAQS regulatory development, and the products of pre-dissemination review can be examined by the Information Quality Review Committee to ensure that they have actually achieved the goals EPA set forth in its Information Quality Guidelines.

5. *Establish an information quality foundation for all CASAC reviews of NAAQS-related work products whose successful performance is documented and independently validated.*

The record shows that EPA did not inform CASAC about information quality principles or the Agency's policies and practices that were established to achieve adherence with these principles. By leaving information quality out of the Charge, never educating CASAC about its meaning and implications, and conducting a multiyear dialogue with CASAC that never broached the subject, EPA staff ensured that CASAC could not and would not ever take account of information quality concerns in its scientific review. The extent to which this failure distorted the committee's policy recommendations cannot be ascertained, but it should be assumed that the committee would have offered different policy advice if information quality had been central to its review of the science.

There is no question that reports prepared by the Information Quality Review Committee would be very useful to CASAC. Committee reviews should be scheduled early and often so that when CASAC convenes to review a NAAQS work product, it has at its disposal a thorough and objective review of the information quality attributes of the scientific information it must examine.

6. *Require that CASAC panel members be recused from the review of their own research or the research of their institutional colleagues, and such individuals should not serve on a CASAC panel in cases where such research is crucial to risk assessment and characterization.*

The CASAC ozone panel, like previous CASAC panels, included a number of members whose primary research interests and activities involve ozone. In some cases, CASAC panel members are the authors or co-authors of studies relevant to the assessment and characterization of health risk, or they are institutional colleagues of such researchers. It is vital that these individuals participate extensively in that portion of the CASAC process which consists of assembling and summarizing scientific data. However, an intellectual conflict of interest arises when these scientists are also asked to review EPA staff work products that interpret their research or judge its quality. It is unreasonable to expect intellectually conflicted CASAC panel members to provide unbiased opinions of EPA staff work products, and the opinions of such panel members cannot reasonably be assumed to be free of self-regard.

In some cases, mere recusal from a portion of CASAC review is not sufficient. The group dynamic of peer review inhibits those panel members who are not conflicted from candidly expressing serious concerns and doubts. Scientists who are authors, co-authors, or institutional colleagues of the handful of scientific studies that are identifiable as crucial should not serve on CASAC at

all. It also would help if EPA did a better job applying its own peer reviewer selection rules to ensure experts selected to serve had open minds. Before the review even began, several members of the CASAC ozone panel were on record supporting major reductions in the ozone NAAQS.

II. Introduction

This RFR is submitted to EPA in accordance with government-wide requirements related to information quality (Information Quality Act 2000; Office of Management and Budget 2002) and procedures established by EPA (U.S. Environmental Protection Agency 2002), concerning certain information disseminated by the Agency in association with its recent proposed rulemaking on the ozone National Ambient Air Quality Standard (NAAQS) (Docket ID EPA-HQ-OAR-2005-0172). Pursuant to these Guidelines, a copy of our RFC is attached. EPA Guidelines recommend that RFRs be submitted within 90 days, but we found that the magnitude of the task was too complex to complete in such a short window.[5]

A. Information Subject to this Request for Reconsideration

The RFC set forth a list of documents that constituted the information subject to the petition. This list included the 3-volume Criteria Document (U.S. Environmental Protection Agency 2006a, 2006b, 2006c), the Staff Paper (U.S. Environmental Protection Agency 2007g, 2007j, 2007k, 2007l), exposure and risk assessments (U.S. Environmental Protection Agency 2007d, 2007e), certain internal memoranda (Brown 2007a; Langstaff 2007), and the preamble to the notice of proposed rulemaking (U.S. Environmental Protection Agency 2007h). All of these documents include influential scientific, technical, statistical and economic information that is subject to the Information Quality Act (Information Quality Act 2000), OMB's government-wide guidelines (Office of Management and Budget 2002), and EPA's agency-specific guidance (U.S. Environmental Protection Agency 2002).

[5] EPA's Information Quality Guidelines are subordinate to the government-wide guidelines issued by the Office of Management and Budget (2002). OMB's guidelines authorize agencies to establish "appropriate time limits in which to resolve such requests for reconsideration," taking account of whether "other agencies may have an interest in the resolution of any administrative appeal " (pp. 8458 and 8459). OMB's guidelines do not authorize agencies to impose artificial deadlines on the submission of such appeals.

In our RFC we also listed EPA's Regulatory Impact Analysis (U.S. Environmental Protection Agency 2007i) as a covered document. However, given the very limited amount of time available between the date it was published (August 2, 2007) and EPA's requirement that a timely RFC meet the deadline for public comments on the proposed rule (October 9, 2007), we were compelled to set priorities and defer this matter until a later date. The RIA does not appear to have been distributed for notice and comment – it is not part of the standard-setting process and we could not locate a relevant Federal Register notice requesting public comment -- so it is not covered by this part of EPA's Information Quality Guidelines, which required that we submit our RFC on or before the deadline for public comments (U.S. Environmental Protection Agency 2002, Section 8.5). Moreover, because the RIA was not a factor in the Administrator's final decision, there is no deadline for timely submission of an RFC with respect to the RIA. Nonetheless, the RIA incorporates scientific information from the documents listed above. Thus, our challenges to the scientific information in the aforementioned documents also apply to the RIA to the extent that the RIA contains materially equivalent or derivative information quality errors.

Our RFC concerned influential scientific, technical, and statistical information contained or referenced in these documents. It did not include material that is strictly policy in nature; such information is excluded from the definition of "information" because it is an expression of values or preferences, and not of facts or data (Office of Management and Budget 2002).[6] Likewise, this RFR also concerns information and not expressions of values or opinion.

B. Affected Party

The National Association of Manufacturers (NAM) is the nation's largest industrial trade association representing small and large manufacturers in every industrial sector and in all 50 states. Headquartered in Washington, D.C., the NAM represents a sector that employs more than 14 million American workers. The NAM's mission is to enhance the competitiveness of manufacturers and

[6] "'Information' means any communication or representation of knowledge such as facts or data, in any medium or form, including textual, numerical, graphic, cartographic, narrative, or audiovisual forms. This definition includes information that an agency disseminates from a web page, but does not include the provision of hyperlinks to information that others disseminate. This definition does not include opinions, where the agency's presentation makes it clear that what is being offered is someone's opinion rather than fact or the agency's views." See Section V(5) at 8460.

improve American living standards by shaping a legislative and regulatory environment conducive to U.S. economic growth.

As the leading voice of manufacturing in the United States, the NAM is deeply concerned that crucial decisions on air pollution control policy reflect the best, unbiased scientific information possible. Our members, and their employees and families, deserve that these important policy decisions be grounded in science.

C. Applicable Error Correction Procedures

Under OMB's government-wide information quality guidelines (Office of Management and Budget 2002), every agency must issue its own implementing guidelines, taking account of its specific needs and characteristics. EPA's Information Quality Guidelines (U.S. Environmental Protection Agency 2002) follow the OMB Guidelines in most material respects. We followed EPA's agency-specific procedures for affected parties in submitting our RFC (Section 8); in particular, we simultaneously submitted the RFC as a public comment on the Notice of Proposed Rulemaking (Section 8.5, page 32).

Independent appeal provisions are set forth in Sections 8.6 and 8.7 of EPA's Information Quality Guidelines. A three-member executive panel consisting of the Science Advisor/AA for the Office of Research and Development (ORD), the Chief Information Officer/AA for OEI, and the Economics Advisor/AA for the Office of Policy, Economics and Innovation (OPEI) normally would investigate the claims decide the appeal after presentation of the issues by the "information owner." In this case, there are two "information owners": the Assistant Administrator for Air and Radiation (OAR) and the Science Advisor/AA for ORD. Because information owners must be recused from the appeal process for it to be plausibly independent, EPA's Science Advisor/AA for ORD cannot serve on the executive panel and must be replaced with an Assistant Administrator other than the AA/OAR, or a Regional Administrator. EPA is required to conduct appeals in a timely manner, and the Agency has decided that 90 days meets this requirement.

D. Relevant Information Quality Principles

Each of the documents that was designated a subject of the RFC is *influential*, as that term is defined in both OMB's and EPA's guidelines. The specific information quality principles at issue are (a) utility, (b) integrity, and (c) objectivity. Objectivity comes in two subspecies: (i) substantive objectivity and (ii) presentational objectivity. Related to but distinct from the twin objectivity principles is a requirement that influential information be transparent and capable of being substantially reproduced. Transparency is essential for

reproducibility, and reproducibility often is necessary for affected parties to be able to detect information quality errors.

1. Failure to adhere to the objectivity standards

In our RFC, we claimed that information within the listed documents did not satisfy the information quality principles of objectivity (both subspecies). In particular, information about ozone health risk is neither substantively objective nor presented in an objective manner. These defects are pervasive and systemic. In some cases they are obvious, and in other cases quite subtle. Because EPA's Regulatory Impact Analysis (RIA) relies on this information as a critical input for the estimation of health benefits, estimates of costs, risks and benefits also are not substantively objective.

We continue to assert that most of the scientific, statistical, and technical information that we challenged via the RFC does not adhere to the information quality standard of objectivity.

2. Failure to adhere to the utility standard

We claimed that because of these systemic and material defects in objectivity, the documents subject to the RFC did not satisfy the *utility* standard. *Utility* requires that information that is disseminated be useful for the purpose to which it was intended. In the case of the RIA, the purpose of the document was to accurately, fully, and clearly inform the public concerning the costs, benefits, distributional consequences, and other effects attributable to a more stringent ozone NAAQS. Pervasive and systemic information quality errors in EPA's risk assessment rendered the Agency's risk and benefit estimates systematically biased, and thus neither valid nor reliable for informing the public. Substantively "accurate, reliable, and unbiased" benefit estimates require, at a minimum, "accurate, reliable, and unbiased" estimates of risk. It is impossible for a benefit estimate to satisfy the substantive objectivity standard if it must rely on crucial information that is materially defective with respect to substantive objectivity. For that reason alone, benefit estimates in the RIA also do not satisfy the substantive objectivity standard, and by failing that standard they can not have utility for their intended purpose of informing the public about the impacts of a revised ozone NAAQS.

The purpose of the Criteria Document and Staff Paper were to accurately, fully, and clearly inform the Administrator concerning the health risks posed by ozone at levels below the current standard, the incidence of health effects resulting from these risks assuming attainment of the current standard, and the change in incidence resulting from alternative, lower standards. Due to EPA's pervasive and systemic failure to adhere to the substantive and presentational

objectivity standards in its risk assessment, it is impossible for the Criteria Document and Staff Paper to have utility for the Administrator so long as he is committed to set the standard in accordance with the criteria established by law. The law does not authorize the Administrator to base his decision on inaccurate scientific information.

The purpose of the preamble to the Notice of Proposed Rulemaking was to articulate, and communicate to the public, the scientific information that the Administrator considered, and the reasoned basis for determining what standard to propose to set. The Administrator has substantial policy discretion provided by law to decide where to set the standard, and the reasoned basis set forth in the preamble explains how the Administrator incorporated the scientific information he was provided. However, this scientific information was fundamentally flawed because it systematically violated the objectivity standards. For that reason, the Administrator's reasoned basis for decision-making almost certainly relies on inaccurate scientific information. In the RFC, petitioners did not challenge the Administrator's reasoned basis for decision-making, for such a challenge is impermissible under both OMB's and EPA's Information Quality Guidelines. Rather, we challenged the scientific and statistical information provided to the Administrator. Nonetheless, it is at least plausible and perhaps highly likely that the Administrator's decision would have been different if he had been provided scientific and statistical information that adhered to applicable information quality principles.

We continue to assert that EPA's failure to adhere to the objectivity standards means that the information still being challenged does not adhere to the utility standard.

III. Information Quality Errors in the Description, Analysis and Presentation of Scientific Information

In our RFC, we noted "[b]ias takes many forms" that "affect the scientific information upon which EPA relies" and "how EPA chooses to utilize this information." We also noted that

> [b]ias per se is not a violation of the information quality standard of objectivity because it is an evitable fact when dealing with uncertain quantities that have to be estimated. However, purposeful bias – the dissemination of information that is known or intended to over- or understate uncertain quantities – is unambiguously a violation of the objectivity standard. Information containing a series of purposeful biases systematically violates the objectivity standard (National Association of Manufacturers 2007, pp. 9-10, emphasis in original).

We alleged that EPA began with a database that was structurally biased by (1) control of development and publication by parties with well-defined risk management objectives; (2) multiple forms of publication bias; (3) systemic methodological error; and (4) peer review in which the assurance and maximization of information quality played no part whatsoever. These problems were severely exacerbated by EPA staff's last-minute submission to the scientific record a critical reanalysis that had never been publicly disclosed, peer reviewed, or subjected to any of the normal procedures of scientific review and validation.

Although peer reviewed scientific information enjoys a presumption of adequate objectivity (Office of Management and Budget 2002), this presumption is a weak one that can be rebutted by a persuasive showing that the information is not in fact objective, or that the peer review on which the presumption rests was deficient in a material respect relevant to information quality. For example, to show that peer review did not assure even adequate objectivity, it is sufficient to show that in their charge the peer reviewers were not asked to evaluate whether the information satisfied applicable information quality principles, or if they were so asked that they failed to fulfill their charge. An agency cannot evade its responsibilities under the Information Quality Act by waving peer review as a talisman or wearing it as an institutional phylactery. For peer review to serve its purposes under information quality guidelines it cannot do so accidentally; it must rigorously apply information quality principles.

Federal agencies, not petitioners for correction or independent research scientists, are subject to the strictures of information quality. This system is not designed to set up the perfect as the enemy of the good, but to deter the government from abusing its unique powers and responsibilities. The scheme granting the government a presumption of adequate objectivity for information that has been peer reviewed has three important and desirable features: it establishes a low hurdle that excludes scientific information that has not endured the minimally invasive rigors of professional scrutiny; it creates an incentive for more-objective information to supplant less-objective information at every reasonable opportunity, thus fostering scientific advancement; and it rewards peer review procedures that explicitly and rigorously evaluate adherence to information quality principles, most notably, the principles of presentational and substantive objectivity. An agency cannot justify a preference for less-objective information because it more conveniently conforms to a policy mission or the risk management preferences of staff or management. To succeed in rebuttal, a petitioner need only show that the information cited approvingly by the agency is materially lacking in objectivity. The petitioner need not show that he "knows" the right answer or that he can point to alternative scientific information that is provably unbiased. Agencies, in short, are not allowed to use bad information

just because it is all they have, or to reject better information because it is not perfect.

A. EPA Begins with a Structurally Biased Scientific Database

In our RFC, we stated that influential scientific information provided to EPA may be biased for several reasons. We discussed three such reasons.

1. Effective control by a party with a risk management objective

We noted that health-effects studies have been funded by government, industry, and sometimes jointly. Because industry has well-defined policy interests, it is often suspected or accused of trying to control research so that it yields desired results. EPA is not shy about highlighting such potential conflicts of interest. In its Response to Comments, for example, four times EPA implies that a study or review might be technically unsound because it was industry-funded.[7] In none of these cases, however, does EPA actually provide evidence suggesting how scientific integrity was in fact compromised. Rather, EPA's approach consists of suggestive condemnation by association – by itself, a material breach of applicable information quality standards because it involves the attempted elucidation of quality distinctions based on criteria other than quality.

Nongovernmental organizations and government agencies like EPA also have well-defined policy interests, and thus they have similar incentives to control research to ensure agreeable outcomes. In EPA's Response to Comments, there is no instance in which EPA implies that a study or review might be technically unsound because it was NGO- or government-funded. Indeed, the list of studies EPA heavily relies on that the Agency itself funded is an extensive one.[8] Nowhere in the Response to Comments, however, does EPA ever

[7] See EPA (2008d, p. 45): p. 5 (Brauer et al. 2007), pp. 21 and 97 (Adams 2006a), and p. 22 (Smith 2007b). EPA praises itself for being "a leader" in examining the so-called "GAM problem," having "funded a special workshop and supported the [Health Effects Institute] in a project to reanalyze dozens of studies to fully investigate this issue" (p. 45). EPA does not mention any other sources of funding, nor does it acknowledge the consequences the Agency would have suffered had it refused to participate.

[8] The list of critical studies funded by EPA but not identified as such in EPA's Response to Comments includes mortality epidemiology (Bell et al. 2005; Bell, McDemott et al. 2004; Bell et al. 2006; Levy et al. 2005); morbidity epidemiology (Korrick et al. 1998; Mortimer et al. 2002; Ross et al. 2002); school absence epidemiology (Chen et al. 2000; Gilliland et al. 2001); and controlled human exposure (McDonnell 1996).

acknowledge the Agency's role or imply that EPA funding might have infiltrated the studies' designs, implementation, results or reporting.

It is because of this asymmetry in EPA's treatment of scientific information that industry routinely funds research through arm's length grants and contracts that insulate researchers from sponsor interference. The extent to which NGOs and government agencies do so is not well documented.[9]

Federal information quality guidelines deal with the problem of sponsor bias two ways. First, they place a high value on full disclosure sufficient to ensure reproducibility. Reproducibility is widely believed to be the best procedural tool for determining whether interference occurred. When a research sponsor declines to make its data available, that which it does disclose may become presumptively suspect. Second, as long as information is capable of being substantially reproduced, information quality principles emphasize quality attributes and not the source of research sponsorship per se. If these principles are adhered to, then biases resulting from sponsor control over research should be rare because they would be detectable.

Many times in EPA's Response to Comments, the Agency attempts to deflect scientific questions raised by many commenters based on the policy preferences of the commenter rather that the scientific merit of the comment.[10] This tactic creates the false perception that the warring interests on both sides are motivated solely by policy disputes and only EPA is motivated by the pursuit of science. This practice is false – EPA staff, managers, and officials all have policy views – and it is anathema to good government because it unfairly stigmatizes the scientific integrity of public commenters generally.

Information quality guidelines require that influential information that an agency proposes to disseminate be capable of being reproduced by a competent and independent third party. This transparency requirement is a precursor step

---

[9] When scientists perform unquestionably independent research funded by industry, it is often then alleged that they skew their work to ensure a steady stream of future research grants. Such allegations can never be disproved because they are not testable. In any case, the identical claim can be made about NGO and government-funded research programs, such as for example EPA's STAR grant program.

[10] For examples in which EPA deflected critical scientific (not policy) comments on the ground that the commenter opposed revising the NAAQS, see, e.g., pp. 12, 14, 15, 19, 30, 37, 40, 58, 75, 77, 104, and 128. For examples in which EPA deflected critical scientific (not policy) comments on the ground that the commenter supported a revised NAAQS below the value selected by the Administrator, see, e.g., pp. 10, 11, 12, 14, 15, 55, 56, 104, and 105.

in the assurance of presentational and substantive objectivity, for it is by checking the government's work that departures from objectivity are most readily detected. If the government could withhold information necessary to enable reproducibility, it could obstruct the public's ability to exercise its legal right to objective information.

In the ozone case, EPA asked for and promptly obtained from Prof. William C. Adams data from several of his controlled human studies.[11] EPA was then able to reproduce a subset of Adams' results, and even to perform a reanalysis of his data to partially test his work for objectivity.[12]

In contrast, EPA or an allied federal agency funded virtually all of the epidemiological studies that Agency staff consider highly influential for estimating human health risks. Moreover, EPA has by law the right to obtain data from researchers who perform Agency-funded research (Office of Management and Budget 1999, Sec. 36). However, we can discern no instance in the ozone review in which EPA has exercised this right. For government-funded research, the Agency's staff is generally satisfied that the information provided in published papers is full and complete.[13]

---

[11] Note that EPA has a pattern of requesting and obtaining data from industry-funded studies: "As in the 1997 risk assessment, EPA obtained individual data from several 6.6-hour O_3 controlled human exposure studies from [Adams]. API, the funding sponsor of the Adams studies, urged EPA to use the data from these studies, particularly the most recent study by Dr. Adams in its health risk assessment in its comments on the draft Staff Paper and draft health risk assessment in January 2006. EPA obtained the individual data used in the health risk assessment directly from the author and explained that the data would be combined with other individual data from the Horstman, Folinsbee, and McDonnell 6.6-hour O_3 studies" (U.S. Environmental Protection Agency 2008d, pp. 97-98, emphasis added, internal references omitted).

[12] EPA requested and obtained only selected data, and proceeded to analyze only this subset. It did not seek to reproduce Adams' analysis. It is noteworthy that in this reanalysis, EPA does not claim that Adams' data or his statistical analysis departed in any manner from the information quality standard of objectivity.

[13] The record shows several cases in which, when questions arose concerning details not reported in refereed articles resulting from EPA-sponsored research, EPA staff simply requested analytic results not included in the published papers from these researchers and cited them as "personal communications" (U.S. Environmental Protection Agency 2006a, pp. 7-179, 177-185, 178-183; 2007g, p. 3-93). Additional data or analytic results provided to EPA via "personal communications" cannot be reproduced by independent third parties, and thus are inherent violations of applicable information quality standards.

We said in our RFC that EPA staff analyzed the scientific record with a policy-driven bias in favor of discovering risk. EPA denies this, but the fact that it regarded its own funded research as inherently trustworthy and industry-funded research as presumptively biased is prima facie evidence that our allegation is in fact correct. A necessary condition for the absence of interpretative bias is persuasive evidence that EPA had in place, and actually followed, a plan for pre-dissemination review that applied the same information quality standards for the review of all scientific information irrespective of how it was funded. Not only did EPA fail to follow such a plan, its Response to Comments reveals that it didn't even have a plan to follow.[14]

A practical consequence of EPA's managerial control over both the scientific record and the development of policy alternatives is the Agency staff appears to have been unable to prevent its policy preferences from influencing its presentation and review of the scientific record. The evidence is overwhelming that these conflicting missions resulted in a systemically biased characterization of the human health effects of ozone.

2. *Publication bias*

In our RFC, we identified three subspecies of publication bias that we believe are present in EPA's scientific record: positive-results bias, outcome-reporting bias, and inferential exaggeration (National Association of Manufacturers 2007, pp. 11-13). Positive-results bias occurs because studies that do not show positive associations are published less frequently, if at all. Outcome-reporting bias occurs when researchers report results with the highest apparent association, a widely observed phenomenon. Sometimes, dozens of models will have been examined but only the handful with the strongest association will be reported (Lumley and Sheppard 2003). Inferential exaggeration occurs when scientists draw (and editors accept) conclusions that are not supported by the data and analysis actually performed.

We noted in our RFC (and EPA did not dispute in its Response to Comments) that positive results bias and outcome-reporting bias are endemic in epidemiological literature. EPA's Response to Comments (U.S. Environmental Protection Agency 2008e, pp. 31-32) says Agency staff "recognized the potential impact of publication bias" in Section 7.1.3.6 of the Criteria Document. EPA staff

[14] See EPA (2008e, p. 150): "EPA's Guidelines for Ensuring and Maximizing the Quality, Objectivity, Utility, and Integrity of Information Disseminated does not require the Agency to discuss, separately, whether the pre-dissemination review actually occurred."

acknowledged the problem of publication bias,[15] set forth a consistent method for addressing the portion of publication bias caused by multiple hypothesis testing,[16] then abandoned this method because it was infeasible and retreated to the default assumption that the problem did not exist.[17] EPA staff also said they would emphasize zero and 1-day lags[18] in tine-series studies and give more weight to primary over secondary analyses.[19] In the Staff Paper, however, EPA staff did exactly the opposite. They relied on the distributed lag models of Mortimer et al. (2002) and Bell, McDermott et al. (2004), defended this reliance in the Response to Comments (U.S. Environmental Protection Agency 2008e, e.g., pp. 32-33, 43-44, 45-48), and made little or no distinction between primary and sensitivity analyses.

Outcome-reporting bias can be documented by comparing the protocols that researchers established prior to beginning work against the analyses that they actually published. Inferential exaggeration can be detected by carefully comparing study results against researchers' inferences to ensure that inferences do not reach beyond what the data and analyses support. To test for outcome-reporting bias, however, EPA staff would have to obtain researchers' data, made the effort to reproduce their work, and explored what alternative results could have been reported using the same data and alternative methods. EPA obtained research data only from industry-sponsored researchers, so it was able to examine outcome-reporting bias only for this subset of research projects. One interpretation of EPA's reanalysis of the Adams (2006) data – Brown (2006, 2007a, 2007b) -- is that Agency staff sought to uncover outcome-reporting bias in the form of statistically significant associations not reported by Adams. There is no

---

[15] "The summary of health effects in this chapter is vulnerable to the errors of publication bias and multiple testing."

[16] "To address multiple hypothesis testing, emphasis will be placed in this chapter on a priori hypotheses."

[17] "As identifying a priori hypotheses is difficult in the majority of the studies, the most common hypotheses will be considered."

[18] "For example, although many studies examined multiple single-day lag models, priority would be given to effects observed at 0- or 1-day lags rather than at longer lags."

[19] "Analyses of multiple model specifications for adjustment of temporal or meteorological trends will be considered sensitivity analyses. Sensitivity analyses shall not be granted the same inferential weight as the original hypothesis-driven analysis…"

public evidence that EPA made any effort at all to uncover outcome-reporting bias among those researchers whose work the Agency itself sponsored.

In our RFC, we had no all-purpose remedy for the problem of positive-results bias because we recognized that it was so hard to detect. With respect to outcome-reporting bias, however, we said

> [f]or each critical study, EPA should determine the extent to which nonpositive outcomes were not reported and include that information in its presentation (National Association of Manufacturers 2007, pp. 12-13).

In its response, EPA gives three reasons for denying our remedy request:

> EPA rejects NAM's contention that it should determine the extent to which nonpositive outcomes were not reported and include that information in its presentation of each critical study. First, there is no evidence to show that researchers are not reporting all results. Second, EPA can not include in its assessment results that were not reported. Third, EPA uses a weight of evidence approach to evaluate evidence that does not depend on a few critical studies (U.S. Environmental Protection Agency 2008e, p. 32).

Each of these reasons is logically irrelevant but revealing.

Addressing these arguments in order, it is generally agreed that the absence of evidence is not the evidence of absence. In this case, the most obvious reason why "there is no evidence to show that researchers are not reporting all results" is that EPA has not looked – except in the case of industry-funded researchers, where it has found none.

Second, EPA is not prohibited from supplementing the scientific record with data and analyses beyond that which was published by original authors, provided that the Agency makes the new data and analyses transparent and reproducible, and subjects them to effective peer review. Indeed, the EPA staff practice of constraining its review to published results encourages and intensifies outcome–reporting bias.

Moreover, if it is true that "EPA can not include in its assessment results that were not reported" in published studies, then it must discard several of its internal memoranda because they rely on unpublished data or nontransparent syntheses of published data.[20] EPA also must discard the unpublished results it

[20] There are many prominent internal memoranda that meet this definition (Cox and Camalier 2006; Langstaff 2006a, 2007; McCluney 2007; McCluney et al. 2006; Rizzo 2005, 2006).

has obtained from Agency-funded researchers.[21] EPA cannot simultaneously rely on unpublished results that it has specifically requested and claim that it is not allowed to rely on unpublished results.

Third, EPA's "weight-of-evidence approach" is inherently noncompliant with EPA's own Information Quality Guidelines. Without researchers "showing their work," these results are not "capable of being substantially reproduced" (U.S. Environmental Protection Agency 2002, p. 47).

In its response, EPA mischaracterizes our complaint as saying that "EPA has not considered publication bias," then proceeds to rebut its mischaracterization rather than our complaint (U.S. Environmental Protection Agency 2008d, p. 31, emphasis added). We made no such claim; our concern was that while EPA may have "considered" it, the Agency's efforts to estimate its magnitude consisted of meta-analyses primarily intended to address uncertainty about the magnitude of effect estimates across studies and strengthen statistical significance.[22]

(a) Outcome-reporting bias

We specifically identified several panel studies on which EPA relies as displaying evidence suggestive of outcome-reporting bias (Gent et al. 2003;

---

[21] Expanding on footnote 13, in the Criteria Document and Staff Paper EPA references unpublished mortality estimates obtained from Bell ("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 communication with attachments to Jee Young Kim]. New Haven, CT: Yale University School of Forestry and Environmental Studies; January 6." (U.S. Environmental Protection Agency 2006a, p. 7-179; 2007m, p. 3-93) and Ito ("Ito, K. (2004) Revised ozone risk estimates for daily mortality and hospitalizations in Detroit, Michigan [personal communication with attachments to Jee Young Kim]. New York, NY: New York University School of Medicine, Nelson Institute of Environmental Medicine; October 31.") (U.S. Environmental Protection Agency 2006a, p. 7-185). EPA reprints the unpublished estimates from Bell in Figure 7-17 of the Criteria Document but not in the Staff Paper. EPA cites the unpublished results obtained from Ito on pages 7-76 and 7-82 of the Criteria Document.

Outputs of statistical analyses delivered by researchers to EPA via personal communication inherently violate EPA's Information Quality Guidelines because they are not reproducible.

[22] In Section I.D.1 on page 14, this phenomenon is first on our list of major types of information quality error. Whether EPA has "discussed" or "considered" something is immaterial; what matters is what the Agency disseminates as authoritative.

Korrick et al. 1998; Mortimer et al. 2002). The extent to which these results are representative of all the models analyzed is not clear, nor is it known how many different models the authors examined before settling on the ones they published.

In its Response to Comments, EPA says "there is no evidence to show that researchers are not reporting all results" (U.S. Environmental Protection Agency 2008e, p. 32). As we noted earlier, EPA's lack of evidence is assured by the staff's having not inquired. Gent et al. (2003) is very tightly written to accommodate the journal's notoriously severe space constraints, but nevertheless it discloses that that a variety of models were examined and not every result was reported. Korrick et al. (1998) acknowledge reporting results only from reduced-form models in which variables that *a priori* they considered important were dropped due to lack of statistical significance. Mortimer et al. (2002) acknowledge examining a wide array of lagged exposure models, and they imply that they analyzed morning PEF values because evening values, which they expected to be diminished from the day's ozone exposure, were not. More generally, it is a ubiquitous practice for journals to publish a (usually small) subset of the analytic work actually performed. EPA's position that "there is no evidence" that what's reported is incomplete is impossible to credit.

(b) Inferential exaggeration

Because they are in the business of conducting research, scientists as a group are predisposed to be cautious about drawing inferences that go beyond their data and analyses. However, because they also have opinions about policy and face other incentives, sometimes they do not follow these professional norms and instead exaggerate the strength or certainty of their results, or the implications of their results for public policy.

In our RFC, we cited Gent et al. (2003) and Mortimer et al. (2002) as examples of refereed papers in which notable inferential exaggeration was present (National Association of Manufacturers 2007, p. 13). Gent et al. (2003) concluded that asthmatic children using maintenance medication are "particularly vulnerable" to ozone, even after controlling for exposure to fine particles, at levels below the current standard. We said that this conclusion went beyond what could be inferred from the reported data and analysis. The language alone is laden with policy judgment; the phrase has no objective scientific meaning.

Mortimer et al. (2002) also concluded that ozone below current standards has adverse effects on asthmatic children (p. 704). In our RFC, however, we said

their conclusion was based on a selective reporting of model results (Id.). The authors performed analyses using an exhaustive set of lag models[23] and reported results from a subset of these analyses, then discussed as representative and meaningful only those results that yielded statistically significant positive effects.

In its response, EPA does not actually rebut our claim that Gent et al. (2003) and Mortimer et al. (2002) engage in inferential exaggeration. With regard to Gent et al. (2003), EPA says only that "there is no reason to believe that the peak O_3 concentrations drive the findings of this study" (U.S. Environmental Protection Agency 2008d, p. 32). EPA ignores both the fact that the study area was in nonattainment and the authors' own statements about their work. The authors say it is a strength of their study design that it "use[d] both the maximum 1-hour average (sensitive to spikes in concentration) and 8-hour average (a measure of short-term, cumulative exposure) to assess daily ambient ozone levels" (p. 1865, emphasis added). In other words, the Gent et al. (2003) study was designed to capture peaks, and to the extent that the study uncovered a few weak concentration-response trends it was peaks that gave these trends away.[24]

(i) The inferences made by Gent et al. (2003) are inconsistent with their own reported results

Examples of significant inferential exaggeration in Gent et al. (2003) are not hard to find:

- "A 4% increase in bronchodilator use was … associated with same-day levels of ozone (51.6-58.8 ppb) (Table 4, model 1)" (p. 1862). The authors do not mention that this odds ratio is barely statistically significant; the 95% confidence interval is 1.00 – 1.09. Also, they do not mention that it is the only statistically significant positive odds ratio among 20 odds ratios reported for five ozone concentration quintiles compared across two exposure scenarios (previous day, same day) and two averaging times (1-hour, 8-hour). There is no statistically significant positive trend for either exposure scenario or either averaging time (range of *p* values: 0.13 to 0.64), a fact they also do not mention. Finally, these results pertain to single-pollutant models,

[23] "Lagged air pollution effects were evaluated using moving averages, unrestricted distributed lags, and polynomial distributed lags." See Mortimer et al. (2002, p. 700).

[24] EPA has said that triangular exposures are more representative of real-world conditions (U.S. Environmental Protection Agency 2007g, p. 6-10).

which would be expected to overstate the effect of ozone if other pollutants (such as fine PM) are positively correlated.

- "In logistic regression models of both ozone and fine particles for children taking maintenance medication, an increased likelihood of respiratory symptoms was associated with levels of ozone on the same day, previous day, or both; and increased bronchodilator use was associated with the highest level of same-day ozone" (p. 1863). Consistently positive trends in same-day 1-hour effects are shown for wheeze and shortness of breath; for chest tightness only in the highest two quintiles, and not at all for persistent cough. Previous day 1-hour effects are statistically significant only for chest tightness. A statistically significant increase in bronchodilator use was reported only for the highest exposure quintile; this increase was reported for only one of ten odds ratios. The authors glossed over these details, focusing instead on the handful of statistically significant effects.
- "In logistic regression models of both ozone and fine particles for children taking maintenance medication, an increased likelihood of respiratory symptoms was associated with levels of ozone on the same day, previous day, or both…" (p. 1864). Their Table 5, however, shows significant positive same-day effects for only two of five symptoms for the 1-hour averaging time, and significantly positive previous-day effects for only one of five symptoms for the 1-hour averaging time. Two of five endpoints displayed positive trends for previous-day 8-hour averaging times; none of the five displayed positive trends for same day 8-hour exposures.
- "In models controlling for ambient fine particle concentration and typically at levels below EPA air quality standards, daily ambient ozone was found to be significantly associated with increased risk of respiratory symptoms and increased use of rescue medication among children with asthma severe enough to require maintenance medication" (p. 1865). Their Table 5, however, shows significant positive trends only for two of five symptoms for same-day 1-hour averaging times, and significant positive trends only for two of five symptoms for previous-day 1-hour averaging times.

Single-pollutant models yield different results than co-pollutant models where exposure to co-pollutants is highly correlated. In the single-pollutant models, nine of the 80 ozone-related odds ratios reported in Table 4 for asthmatic children on medication are statistically significant. In all cases, it is the previous day's ozone for which the positive trend is statistically significant. No same-day

trends were observed. In the co-pollutant models (Table 5), there are 80 odds-ratios reported and only seven are statistically significant. Two of the three consistently positive trends appear in same-day ozone exposures. This pattern might make sense, but Gent et al. (2003) do not try to explain why. Instead, they reach conclusions identical to their premises but not supported by their own data and analysis.

In our RFC (p. 14), we faulted the EPA staff's reliance on Gent et al. (2003) in part because the authors relied heavily on self-reported symptoms. In its response, EPA says our claim is factually "incorrect." The Agency writes, "In addition to respiratory symptoms, Gent et al. also observed an association between O_3 and rescue medication use, which is an objective measure" ((U.S. Environmental Protection Agency 2008d, p. 48).

EPA is correct that bronchodilator use is a more objective indication that symptoms have led subjects to take action, and that a decision to take action is a better threshold indicator of potential adversity than are changes in FEV_1 or FVC too small for subjects to even notice. However, the results reported by Gent et al. (2003) are inconsistent with their reported (and oft-repeated) claim that they actually found an association between O_3 concentration and bronchodilator use. Their Table 4 shows odds ratios for each quintile (except the baseline) for 1- and 8-hour exposures to previous- and same-day concentrations: four single-pollutant analyses containing 16 odds ratios. Of these 16 odds ratios, one is weak and marginally statistically significant (OR, 1.04; CI, 1.00-1.09). It is the middle exposure quintile for same-day 1-hour exposures; there is no theory we know that would predict this result. We can even discard statistical significance for the individual odds ratios and look for positive trends. Tests performed by Gent et al. (2003) are show a range of p values for positive trends ranging from 0.13 to 0.64.

The claim by Gent et al. (2003) that they found an association between ozone concentration and bronchodilator use is an excellent example of inferential exaggeration, even if it's one that EPA has pointed out to us rather than *vice versa*. Indeed, EPA staff seem to agree that the association claimed by Gent and coworkers does not actually exist, for they have abandoned epidemiology in favor of a policy-driven constraint that defines any increased symptoms, whether dose-related or not, as evidence of an ozone effect unless they can be proved to have another cause:

> Regardless [of what Gent et al. (2003) actually found], EPA deems respiratory symptoms to be a valuable health outcome, which considered in conjunction with various other more "objective" measures, allows a

> more complete depiction of the potential respiratory health effects of pollutants (U.S. Environmental Protection Agency 2008d, p. 48):

Inferential exaggeration is a tool for disguising researchers' policy views under a cloak of science. Thus, when Gent et al. (2003) say that "current standards do not protect these more vulnerable members of the population" (p. 1859), they are expressing their policy view that the primary ozone NAAQS ought to be lowered because science cannot define either *vulnerability* or *adequate protection*. They are, of course, entitled to that opinion. Scientific evidence might help inform their opinion, but it also might not, as policy views are often driven by values, and values influence how much evidence is considered sufficient to either act or stay put.

Scientific peer reviewers often detect and excise inferential exaggeration, but it is entirely plausible that it went undetected in this case. Gent et al. (2003) was accompanied by an editorial that repeated the authors' inferential exaggerations as if they were the same as the authors' research results:

> In the group using maintenance medication, the level of ozone exposure was significantly associated with worsening of symptoms and an increase in the use of rescue medication (Thurston and Bates 2004, p. 1915).

Despite its scientific limitations, Thurston and Bates interpret the evidence reported by Gent et al. (2003) to mean "there is no reason to doubt that ozone exposure is a cause of asthma exacerbations"(p. 1916). When scientists express certainty, it is a powerful hint that they are not talking about science. Their willingness to throw scientific caution to the winds is entirely consistent with holding the unshakeable policy conviction that the ozone standard ought to be lowered irrespective of the scientific evidence. The EPA staff decision to "deem" respiratory symptoms as adverse irrespective of severity or reversibility has a foundation in the Thurston and Bates editorial.[25]

---

[25] "But regardless of the role of air pollution as a contributing factor to the prevalence of asthma, the study by Gent et al. and others like it indicate that the increasing numbers of children with asthma represent an expanding pool of children at risk for respiratory symptoms caused by air pollution, and by ozone in particular")…. Of the many triggers of asthma in the environment, air pollution is one of the few that can be legislated and regulated" (Thurston and Bates 2004, p. 1916, emphasis added). EPA relies heavily on scientific studies performed by Thurston and Bates (there are at least a dozen references in the Staff Paper in which one or both are co-authors), thus raising an obvious question: Where does their scientific analysis end and their policy advocacy begin?

EPA has an obligation under information quality guidelines to conduct sufficient pre-dissemination review to ensure that the information it relies on satisfies information quality standards. That includes detecting inferential exaggeration in scientific papers, limiting the inferences it draws from scientific papers that engage in inferential exaggeration, and subjecting those papers to more rigorous review. In the Staff Paper, EPA includes Gent et al. (2003) within a list of studies which it says "have reported fairly robust associations between ambient O_3 concentrations and daily symptoms/asthma medication use, even after adjustment for co-pollutants" (U.S. Environmental Protection Agency 2007m, p. 3-11). This represents the conclusions Gent et al. (2003) reached, but it does not accurately describe the actual study. It is difficult to honestly infer robustness from a set of exploratory analyses in which all of the odds ratios reported are low and only a handful of them are statistically significant.[26]

(ii) The inferences made by Mortimer et al. (2002) are inconsistent with their own reported results

A similar story can be told with respect to Mortimer et al. (2002). The authors relied on a substandard research design consisting of unvalidated opinion about symptoms obtained from caregivers,[27] limited researcher contact,[28] and the use of diaries to record caregiver-reported data.[29] Response rates were well below the 80% minimum normally required to assume that the presence of

[26] We point out in Section III.A.2(d)(i) that unvalidated self- and caregiver reported data recorded in diaries have been shown to be unreliable. In Section III.C.5(c) beginning on page 102, we note that EPA frequently uses the term "robust" and its variants to describe the consistency of the scientific evidence, but never defines the term.

[27] "Study children had either: 1) parental report of physician-diagnosed asthma and symptoms in the past 12 months or 2) respiratory symptoms consistent with asthma, such as cough, wheezing or shortness of breath, that lasted w6 weeks during the previous year, together with increased symptoms with exercise or cold air exposure or a family history of asthma" (Mortimer et al. 2002, p. 700).

[28] Researcher contact consisted of "an in-person baseline interview, a home survey, three brief telephone follow-up interviews at three-month intervals, and two-week peak expiratory flow rate (PEFR) and symptom diaries after the baseline interview and prior to each follow-up interview" (Mortimer et al. 2002, p. 700).

[29] Diaries were filled out only for the two-week intervals prior to a scheduled "brief telephone follow-up interview." See (Mortimer et al. 2002, p. 700). In Section **Error! Reference source not found.** beginning on page **Error! Bookmark not defined.**, we remind EPA that data from self-administered PEFR meters recorded in diaries have been shown to be unreliable and include manufactured data.

fatal nonresponse bias is not a material defect,[30] and the standard errors calculated by the authors assume without evidence that nonresponse bias is not present.

The authors found no evening effects from ozone that could be measured using self-administered PEF devices. Given the toxicological evidence and chamber exposure studies, one would have expected that evening PEF values exceed morning readings if ozone exposure during the day was causing respiratory effects. But Mortimer et al. did not find elevated evening PEF values. They speculate about how ozone exposure might cause morning but not evening decrements in PEF, and EPA has obligingly amplified their ruminations in the Staff Paper (U.S. Environmental Protection Agency 2007g, p. 3-10).

At this point, all further analysis in Mortimer et al. (2002) should have been understood (both by the researchers and by EPA) to be entirely exploratory. That is especially so given that the claimed statistically significant decrements in morning PEF are all less than 1%. This is a small fraction of the variance in inter-maneuver performance.[31]

EPA admits that Mortimer et al. (2002) chose this multiday lag structure only after it became clear that single-day lags were not going to yield statistically significant positive effects:

> Examination of these single lag day effects led to the consideration of a multiday lag period of 1 to 5 days in the case of PEF and 1 to 4 days in the case of respiratory symptoms to estimate the cumulative effect of O_3 (U.S. Environmental Protection Agency 2008d, p. 32, emphasis added).

With enough statistical effort, Mortimer et al. (2002) were able to unearth a few lags that displayed small statistically significant positive effects, but only buried amidst a huge number of others that were not. For example, none of the

---

[30] Response rates for federally-sponsored surveys with response rates below 80% require an analysis of nonresponse bias simply to be approved under the Paperwork Reduction Act. See (Office of Management and Budget 2006, p. 16). Mortimer et al. (2002, p. 700) report that "approximately 60% of the children returned a diary for each of the four visits." However, they do not reveal the response rate for all four visits, and they do not report the results of any analysis of nonresponse bias. For more about the problem of nonresponse bias as a fatal information quality defect, see section III.A.2(d)(v).

[31] See section III.A.2(d)(i) for a discussion of inter-maneuverinter-maneuver variation in spirometry that is consistently discarded by researchers using these technologies.

single day lags reached statistical significance. The incidence of ≥ 10% declines in PEF was statistically significant for a 4-day lag and a 5-day moving average, but not for lags of zero, 1, 2, 3, or 5 days. The increased incidence of self-reported symptoms was statistically significant for lags of 2 and 4 days, but not lags of 1, 3, 5 or 6 days. See Mortimer et al. (2002, p. 702, Table 2).

Reviewing EPA's Response to Comments led us to take a second look at another vitally important epidemiological study, the time-series analysis by (Bell, McDermott et al. 2004). This time we noticed how frequently the authors described their work as *exploratory*.[32] Despite this cautious adherence to scientific standards in the description of their analyses, for some reason in their conclusions they said their results provided "strong evidence of an association" and implied that this association was both causal and large.[33] They did this even though the used exploratory methods, found weal effects, and could only speculate that the mechanism for ozone exposure causing death "may differ" from the mechanism whereby it caused minor respiratory effects.[34]

---

[32] "Distributed-lag models are appropriate for estimating relative rates of mortality associated with exposure to pollution levels during several previous days, thus allowing more flexibility for exploring the lag between exposure and death than single-lag models. At the second stage, we use hierarchical models to combine the relative rate estimates obtained from the community-specific distributed-lag models to produce a national average estimate. With this 2-stage model, variation across communities in the short-term effects of ozone can be explored and an effect estimated for the nation." (Bell, McDermott et al. 2004, p. 2373, emphasis added and internal citations omitted); "We explored whether the association between ozone and mortality was modified by the long-term average of $PM_{2.5}$ (PM with an aerodynamic diameter less than 2.5 μm) by performing a weighted second-stage linear regression with the community-specific estimate of ozone's effect on mortality as the dependent variable and the long-term $PM_{2.5}$ average as the independent variable. No association was observed" (p. 2376, emphasis added).

[33] "This multisite time-series study of 95 large US urban communities throughout a 14-year period provides strong evidence of an association between mortality and short-term exposure to ozone"; "The results indicate a substantial health burden from ozone pollution" (Bell, McDermott et al. 2004, p. 2376, emphasis added).

[34] "Although the temporal dynamics of the underlying processes linking ozone exposure to increased mortality may differ from those of the inflammatory response, inflammation has been postulated as having a central role in the increased mortality and morbidity associated with ozone "(Bell, McDermott et al. 2004, p. 2377, emphasis added).

(c) Methodological error

In our RFC, we said examples of methodological error can be found in several studies on which EPA heavily relies (National Association of Manufacturers 2007, pp. 14-16). As examples, we mentioned several studies of respiratory symptoms; here, we repeat our examples in the order in which we presented them and discuss EPA's response.

We clearly defined what we meant by a "material effect" – one large enough that it impeded the Administrator's ability from making his decision based on an accurate scientific record (National Association of Manufacturers 2007, p. 7). EPA says it "does not agree that methodological errors exist in these studies that are 'so severe that they have a material effect on utility, particularly for regulatory decision-making'" (U.S. Environmental Protection Agency 2008e, p. 47). However, EPA does not provide much more in the way of rebuttal, as if this is a dispute about policy rather than the application of information quality principles that EPA itself has enunciated and claims to uphold.

- Repeated statistical tests are performed without apparent regard for the resulting increase in the rate of false positives (Korrick et al. 1998; Mortimer et al. 2002).

EPA replies generally that it "conducted a rigorous assessment of potential methodological error in epidemiologic analyses" in section 7.1.3 of the Criteria Document (U.S. Environmental Protection Agency 2008d, p. 47). Yet, EPA's presentation in the Criteria Document discusses, but does not actually address, any of the issues we raised. Neither Korrick et al. (1998) nor Mortimer et al. (2002) come up in this discussion despite the heavy weight EPA places on them in its weight of evidence review.

EPA acknowledges that Korrick et al. (1998) performed repeated statistical tests, but says

> these hypotheses can be divided into confirmatory vs exploratory hypotheses. The main confirmatory hypothesis is whether O_3 concentrations are associated with pulmonary function (U.S. Environmental Protection Agency 2008d, p. 47).

EPA incorrectly equates exploratory data analysis with sensitivity analysis, a tool for evaluating the extent to which the results of a confirmatory data analysis are robust with respect to model specification and other assumptions. Properly understood, exploratory analysis is "detective work" undertaken to get clues

about what a data set might have to say (Tukey 1977).[35] In contrast, Korrick et al. (1998) subjected their data set to numerous statistical techniques in search for the "best" (i.e., strongest) evidence of concentration-response. EPA staff, in turn, rely on "best" statistical results to push the ozone risk envelope outward, and the Administrator relied on the staff's opinion as if were objective (U.S. Environmental Protection Agency 2007h). EPA has insisted that a Bonferroni correction is appropriate for accounting for multiple comparisons in controlled human studies without resulting in excessive Type II error (Brown 2007a). If that correction were applied to the results reported by Korrick et al. (1998), none of the associations reported would have been statistically significant.

In EPA's rebuttal, the Agency staff say that if they cannot find statistical significance they will look for "patterns":

> EPA notes that while statistical significance (i.e., confidence intervals) is considered in the evaluation of the scientific evidence, EPA has emphasized the importance of examining the pattern of results across various studies and not focusing solely on statistical significance as a criterion (U.S. Environmental Protection Agency 2008e, p. 33).

This is more evidence of the Iron Law of EPA Staff Ozone Risk Assessment and Characterization. A dispassionate review of EPA's scientific record shows that in every case where an association is statistically significant, EPA staff interpret that association as evidence of a causal relationship between ozone and health effects. In every case in which an association is not statistically significant, EPA concludes that it nevertheless supports a causal inference. In every case in which a "pattern" of effects is found, that "pattern" also is evidence supporting an inference that a causal relationship exists. Where none of these conditions occur, the study was poorly designed or insufficiently powerful to detect an effect.

- Statistically significant but biologically implausible lags are reported (Mortimer et al. 2002).

In its response, EPA correctly says that it is a reasonable technique in exploratory data analysis to compute and report a wide variety of lags when a biological basis for predicting a specific lag is lacking. However, Mortimer et al. (2002) do not describe their work as exploratory. Rather, they imply that their (statistically significant) moving-average distributed lag model yields results that

[35] In the Criteria Document, EPA confuses sensitivity analysis (performed to illustrate the significance of uncertainty) with exploratory data analysis (the use of statistical tests for which there is no underlying theory to generate new hypotheses) (U.S. Environmental Protection Agency 2006a, p. 7-19).

more accurately describe the pattern of effects than do estimates obtained from (nonsignificant, mixed and oftentimes nonpositive) individual day lag models.[36] They make sweeping, unsupported generalizations about the practical importance of their results.[37]

The EPA staff is clearly smitten by Mortimer et al. (2002). The Criteria Document glosses over the absence of a nonresponse bias analysis despite the low response rate and the absence of evening effects; credits the authors for having "discussed biological mechanisms for delayed effects on pulmonary function" that might support their preferred model, but without characterizing these explanations as speculative; and reports the results of an extensive but nonreproducible additional analysis based on data not reported in the published paper. The Criteria Document repeats the most statistically significant findings reported in Mortimer et al. (2002), but without even the authors' own understated caveats (U.S. Environmental Protection Agency 2006a, pp. 7-43 to 47-46).

EPA's Response to Comments also repeats the post hoc rationalization for moving average distributed lag models using morning-only effects: it is "consistent with the understanding that the development of asthma exacerbation through an inflammatory mechanism would occur over time, with symptoms manifested hours after the exposure period" (U.S. Environmental Protection Agency 2008e, p. 48). But the stated purpose of Mortimer et al. (2002) was to "estimat[e] individual mean effects and individual change over time as well as population mean effects over the entire study period" using "methods [that] require no assumptions about stability of population characteristics over time…" (p. 699). The absence of evening effects was unexpected and the published article offers no explanation why the presumed inflammatory mechanism responsible for morning effects would shut itself off in the evening.

- Single rather than multipollutant models are emphasized (Korrick et al. 1998; Mortimer et al. 2002).

---

[36] "Findings in these USA inner-city asthmatic children are comparable to findings reported elsewhere, suggesting the magnitude of the air pollution-related effect on asthma morbidity is not substantially greater in this population in relation to more socioeconomically diverse groups of asthmatic children" (Mortimer et al. 2002, p. 704).

[37] "In conclusion, summer-time air pollution is associated with increased asthma morbidity and decreased pulmonary function among inner-city children with asthma in the USA. These findings from generalized estimating equations and mixed models support previously published reports from time-series analysis, and those reported from less urban populations" (Mortimer et al. 2002, p. 705).

Our complaint was that EPA relied on single- rather than multi-pollutant models, thus overstating the effect of ozone even if all other considerations were ignored. EPA does not directly respond to this specific error claim (U.S. Environmental Protection Agency 2008d, pp. 47-48), but it does respond to related questions elsewhere in the Response to Comments. EPA's first defense is that Agency staff "include[d] and discuss[ed] results from both single- and multipollutant models when available" and "rigorously and thoroughly evaluated the potential for confounding," but concluded that "the inclusion of copollutants into the models did not substantially affect O_3 risk estimates" and "that effects of O_3 on various health outcomes were robust and independent of the effects of other copollutants" (U.S. Environmental Protection Agency 2008e, pp. 44-45, emphasis added). EPA's second defense is opposite the first: multi-pollutant models were troublesome because they resulted in "reduced stability of the O_3 coefficient estimate in such models (p. 88). This is another example of the Iron Law of EPA Staff Ozone Risk Assessment and Characterization in action: multipollutant models with equivocal results support the current location of the ozone risk envelope; multipollutant models that conflict with that risk envelope are discarded.

- Known confounders are inadequately controlled (Gent et al. 2003; Korrick et al. 1998)

EPA does not directly respond to this specific error claim (U.S. Environmental Protection Agency 2008d, pp. 47-48), but it does respond to related questions elsewhere in the Response to Comments. EPA says it is satisfied with its review of the problems of confounding in epidemiological studies (p. 40), but this satisfaction appears to be focused on co-pollutants rather than non-air pollution confounders. EPA also says that factors such as cockroach and dust mite allergens are unlikely to be genuine confounders in time-series studies because they are not "temporally correlated with O_3" and that they "do not vary from day to day as do ambient O_3 concentrations" (p. 42), though it does not mention that ambient ozone concentrations are not found indoors.

We identified Gent et al. (2003) and Korrick et al. (1998) as studies where control for confounding was inadequate. The most obvious problem is that both did not control for relative humidity, and humidity is clearly associated with pulmonary function changes (Ross et al. 2002). It also seems plausible that the lowest ozone levels occurred on days when humidity was relatively low. The authors also did not obtain data from subjects concerning bronchodilator use, yet there should be no question that the use of such medication will significantly affect respiratory function indicators – that's what they are supposed to do. Pollen is a known cause of allergic rhinitis and asthmatic symptoms, and different pollens have been associated positively -- or inversely -- with

pulmonary function, with magnitude much greater than those for ozone (Ross et al. 2002).[38] The Criteria Document discusses pollen as a confounder only in the context of school absenteeism (U.S. Environmental Protection Agency 2006a, p. 7-58), even though the primary author of Ross et al. (2002) was an EPA employee. Gent et al. (2003), Korrick et al. (1998), and Mortimer et al. (2002) do not attempt to control for pollen.

Controlling for confounding tends to reduce estimated effects. Thus, it is generally incompatible with the Iron Law of EPA Staff Ozone Risk Assessment and Characterization. Ineffective control for confounding results in little change in effect estimates and yields equivocal results that, under thee Principles, support the prevailing location of the ozone risk envelope. Control for inappropriate confounders, such as PM_{10} rather than $PM_{2.5}$ (Bell, McDermott et al. 2004), also will not materially change effect estimates and give the false impression that confounding is not a problem.

If EPA took substantive and presentational objectivity seriously, it would compile a balanced portfolio of causes and risk factors for each major health effect of interest instead of trying to force a causal relationship to air pollution in every case. If EPA had done this with respect to asthma, for example, it would have noticed that its prevalence is rising at the same time that air pollution is falling. Any contribution air pollution might be making thus must be declining, and something else of much greater public health significance is going on. Several recent papers have estimated much larger associations between asthma prevalence and colonization by the gastric bacterium *Heliobacter pylori* in the human gut. Using data from NHANES III, a well-known representative population sample, Chen and Blaser (2007) found that the presence of cagA+ *H. pylori* strains was inversely related to ever having asthma (OR, 0.79; 95% CI, 0.63-0.99), supporting the hypothesis that the lack of normal acquisition ore retention

[38] "O_3 was correlated positively with temperature, as has been frequently observed. It was also correlated positively with grass pollen count, *Curvularia*, and *Drechslera*, and correlated negatively with ragweed pollen count and several mold genera (*Alternaria, Cladosporium, Epicoccurn*). (These correlations were based primarily on seasonal variability and not physical relationships, as with O_3 and temperature.)" Associations with pulmonary function are subject to the limitations and caveats mentioned elsewhere with respect to spirometry, but ozone, temperature and pollen count correlations are not. Reported effects per 20 ppb ozone were 2.3% PEFR decrements in the morning and 2.6% PER decrements in the evening, both before adjustment for aeroallergens. The reported effects from pollen and spores ranged from a PEFR decrement of 6.8% to a PEFR improvement of 31%. See Table 3.

of *H. pylori* is associated with childhood asthma and allergy.[39] A follow-up paper examining very young children obtained even stronger results: the odds ratio for onset of asthma among children under five years of age who had acquired *H. pylori* was 0.58; 95% CI, 0.38–0.88) (Chen and Blaser 2008). As long as EPA persists in trying to link every conceivable health effect to air pollution, it cannot break free of the Iron Law of EPA Staff Ozone Risk Assessment and Characterization.

(d) Pulmonary function testing

Pulmonary function testing is crucial to many of the studies on which EPA relies. In our RFC, we highlighted the information quality problem that these techniques have for non-clinical purposes:

> To obtain reliable data, the procedure requires both training of the person administering the test and practice by the subject, who also must be willing and able to cooperate. Because of the learning effect, multiple tests are necessary to obtain clinically reliable information (National Association of Manufacturers 2007, p. 14).

In its response, EPA summarizes our complaint accurately but answers one we didn't raise concerning reproducibility across devices. We followed that thread, however, and discovered that information quality defects in this body of research are much worse than we originally thought.

(i) Clinically useful pulmonary function tests have inherent information quality limitations that become defects when used in air pollution epidemiology

Pulmonary function tests used to estimate the effects of ozone rely on guidelines published by the American Thoracic Society (ATS) (Miller, Hankinson et al. 2005).[40] These guidelines show that successful testing in clinical settings depends on a combination of factors including the skill of the technician administering the test, the environmental conditions under which the test is

---

[39] For inverse effects, odds ratios are expected to be less than 1. Effects reported to be statistically significant must have upper confidence intervals less than 1.

[40] The most recent editions of the ATS guidance were published in 2005 (MacIntyre et al. 2005; Miller, Crapo et al. 2005; Miller, Hankinson et al. 2005; Pellegrino et al. 2005; Wanger et al. 2005), but on the margins relevant to this discussion earlier ATS guidance is not materially different.

performed, and the level of training and coaching subjects receive. The definition of an "acceptable" test is complex and subject to technician judgment.[41]

After three "acceptable" maneuvers, the two largest FVC and FEV_1 values each must be within 150 mL, or about 3%. Additional maneuvers up to eight can be performed to achieve this error bound, and the technician must "[s]ave, at a minimum, the three satisfactory maneuvers" (p. 325, Table 5). Data should not be discarded solely on the basis for poor repeatability, and the largest values of FVC and FEV_1 should be recorded. If a mid-expiratory flow is taken, it must be measured with an accuracy of ± 0.5% (p. 326).Children present special complications, so ATS recommends that technicians administering tests to children be specially trained.[42]

The ATS guidelines were intended for use in clinical settings where the purpose is to diagnose disease. Reflecting the guidelines' complexity, there is some evidence suggesting that physicians and nurses who administer pulmonary function tests in primary health care settings do not do so very

---

[41] Miller, Hankinson, et al. (2005, p. 325): "The acceptability criteria are a satisfactory start of test and a satisfactory [end of test], i.e. a plateau in the volume–time curve. In addition, the technician should observe that the subject understood the instructions and performed the manoeuvre with a maximum inspiration, a good start, a smooth continuous exhalation and maximal effort. The following conditions must also be met: 1) without an unsatisfactory start of expiration, characterised by excessive hesitation or false start extrapolated volume or EV .5% of FVC or 0.150 L, whichever is greater; 2) without coughing during the first second of the manoeuvre, thereby affecting the measured FEV_1 value, or any other cough that, in the technician's judgment, interferes with the measurement of accurate results; 3) without early termination of expiration; 4) without a Valsalva manoeuvre (glottis closure) or hesitation during the manoeuvre that causes a cessation of airflow, which precludes accurate measurement of FEV_1 or FVC; 5) without a leak; 6) without an obstructed mouthpiece (e.g. obstruction due to the tongue being placed in front of the mouthpiece, or teeth in front of the mouthpiece, or mouthpiece deformation due to biting); and 7) without evidence of an extra breath being taken during the manoeuvre."

[42] "A bright, pleasant atmosphere, including age-appropriate toys, reading material and art, is important in making children feel at ease. Encouragement, detailed but simple instructions, lack of intimidation and visual feedback in the teaching are important in helping children to perform the manoeuvre. Even if unsuccessful at the first session, children will learn to be less intimidated and may perform far better in a subsequent session. Testing children in ''adult'' laboratories, where no effort is made to cater for the specific needs of the younger subjects, is to be discouraged" (Miller, Hankinson et al. 2005, pp. 323-324).

competently. In a randomized prospective study performed in New Zealand, Eaton et al. (1999) found that even after training only 13.5% of patients produced spirometric data that met ATS standards. In the field epidemiology studies on which EPA relies, investigators do not disclose what they did to achieve "acceptable" data.

ATS guidelines also cover the interpretation of spirometric abnormalities (Pellegrino et al. 2005, p. 957, Table 6). The least severe category is "mild" and encompasses all FEV_1 values greater than 70% of predicted; there is no "normal" category.[43] ATS counsels against overinterpreting small changes because intra-personal variability is high:

> It is more likely that a real change has occurred when more than two measurements are performed over time… [S]ignificant changes, whether statistical or biological, vary by parameter, time period and the type of patient. When there are only two tests available to evaluate change, the large variability necessitates relatively large changes to be confident that a significant change has in fact occurred. Thus, in subjects with relatively ''normal'' lung function, year-to-year changes in FEV_1 over 1 yr should exceed 15% before confidence can be given to the opinion that a clinically meaningful change has occurred.[44]

Overinterpreting small changes is a consistent feature of the panel studies the EPA staff relies on to support its inference that ozone concentrations below the 1997 NAAQS cause morbidity. To give just one prominent example, Korrick et al. (1998) reported (and EPA gave considerable weight to) group mean FEV_1 decrements of 2.6% per 50 ppb ozone. Assuming linearity, that's a decrement of 0.4% over the 15 ppb difference between the 1997 and 2007 primary ozone standards. This is less than 10% of the variation in FEV_1 that ATS judges to be clinically meaningful, and is after discarding inter-maneuver variance.

---

[43] None of the studies EPA relies on obtained FEV_1 decrements outside of this category.

[44] See Pellegrino et al. (2005, p. 961, Table 12). Also judged by ATS to be not "clinically meaningful" in normal subjects: (1) within-day changes less than 5% in FVC or FEV_1, and (2) weekly changes less than 11% and 12%, respectively.

(ii) Information quality defects associated with investigator bias

The ATS guidelines specifically directs technicians performing spirometric measurements to coach their patients to ensure best performance.[45] In a clinical setting where the purpose is diagnosis, this is a manageable concern. In a research setting, however, coaching imparts unknown bias to the data. Epidemiologists are not blind to either the hypotheses they are testing or the identity of their subjects, and even if they are scrupulous in their efforts to be unbiased in test administration, test results can be expected to vary across technicians with different coaching skill.

(iii) Information quality defects associated with the use of diaries

EPA's account of the PEFR measurements in Mortimer et al. (2002) is very positive, specifically noting that the National Cooperative Inner-City Asthma Study "used standard protocols that included instructing caretakers of the subjects to record symptoms in the daily diary by observing or asking the child" (U.S. Environmental Protection Agency 2007m, p. 3-11).

In our RFC, we said information quality problems associated with diaries such as these posed special problems. We noted that at least one very high quality study had been performed to ascertain the reliability of data recorded by children's caregivers. Serendipitously, the information that caregivers were supposed to record were the results of spirometric monitoring.

Kamps et al. (2001) studied 40 asthmatic children aged 5-16 years to ascertain the validity of PEFR data self-reported over four weeks by patients and their parents. Data were obtained by diary and, unbeknownst to subjects, microchip memory recorders within the PEFR meters. The simultaneous collection of self-reported and automated data from the same individuals over the same time period provided a powerful test of validity and reliability. Stated compliance with the data collection protocol was 96%, but actual compliance averaged 77%, declining significantly over the course of the study. For 12.5% of the subjects, actual compliance was less than 50%, meaning that nonresponse was systematic and not random. These declines were statistically significant using repeated measures ANOVA. Data were correctly recorded only about half

[45] "The subject should be prompted to ''blast,' not just 'blow,' the air from their lungs, and then he/she should be encouraged to fully exhale. Throughout the manoeuvre, enthusiastic coaching of the subject using appropriate body language and phrases, such as 'keep going', is required" (Miller, Hankinson et al. 2005, p. 323).

of the time; incorrectly recorded about 30% of the time; missing about 6% of the time; and invented between one–eighth and one-fourth of the time. Self-reported data were biased toward understating electronically measured respiratory performance.[46]

Kamps et al. (2001) concluded that self-reported PEFR data were unreliable, and that electronic meters should be used instead of diaries. Similar data obtained by ozone researchers who relied on diaries might have been much more reliable than what Kamps and coworkers found, but the information they report does not offer much comfort. Children enrolled in the cohort examined by Mortimer et al. (2002) were "4–9 yrs old and resided in inner-city neighbourhoods in which the income of ≥ 30% of residents was below the federal poverty level" (p. 700). Caregivers are not well described; the authors only say "children and their parents were recruited." Data were supposed to be recorded in diaries, but the response rate was no greater than 60% for a single round -- worse than the response rate reported by Kamps et al. for their entire study.[47]

A more recent study of adult asthmatics being actively treated yielded similar results (Chowienczyk et al. 1994)618. Diaries contained only 70% of the expected number of records, and 26% of recorded entries were invented or mistimed:

> The rationale behind inventing data or entering data retrospectively may be patients' reluctance to admit poor record keeping. The most striking example to support this is the patient who performed 54 forced expirations in three hours on one day and entered these data retrospectively for the previous six days.

---

[46] This degree of nonresponse, and the problem of manufactured data, likely would have prevented EPA from obtaining permission to collect such data or sponsor its collection under the Paperwork Reduction Act. See Section III.A.2(d)(v) beginning on page 57.

The problem of systematic understating actual values presents additional problems. It means some subjects wanted to be perceived as worse off than they actually were, and that means they cannot be trusted to produce valid data even if they could be persuaded to record it correctly.

[47] Mortimer et al. (2002, p. 700) say "[a]pproximately 60% of the children returned a diary for each of the four visits," but they do not report how many children returned diaries for all of the study period.