

August 15, 2016

Air Quality Management Plan (AQMP) Staff
South Coast Air Quality Management District
Diamond Bar, CA
aqmp@aqmd.gov

Re: Public Comments on Draft 2016 AQMP

Ladies and Gentlemen,

As a follow-up to my unanswered public comments on the 2012 AQMP, I am submitting these public comments on the Draft 2016 AQMP. In particular, I protest the efforts of the SCAQMD staff to try to make the air pollution research pig's ear that uses unreliable, sometimes risible claims of deaths from small particle air pollution (PM2.5) into a silky evidence that justifies more expensive and onerous air regulations.

I think it deceitful that the South Coast would allow small associations as evidence of their claims of thousands of deaths annually in the South Coast area from research that shows small associations of deaths from small particle air pollution. Such claims are riven with deceit. So many papers used by the SCAQMD staff contain small associations and confidence intervals that cannot support the death claims.

I have attached to this cover letter the following items as detailed criticism of the PM2.5 premature death claims made in the Draft 2016 AQMP.

1. My 15-page January 19, 2016 letter to Mr. Henry A. Roman of Industrial Economics, Inc., which takes down the laughable claims of the 2015 Thurston EHP paper that has only small associations (not proof of lethality or toxicity at all) and confidence intervals that cross 1.0 that fail to prove of any death effect at all from small particles air pollution. Dr. Thurston admits the weakness of his evidence in the abstract. He can't make his lack of evidence of deaths from air pollution go away. Data torturing and harvesting noise in the variability of death rates can't fix his problem.
2. The 53-page October 4, 2012 sworn declaration of US EPA senior research scientist Robert B. Devlin in human exposure experiments with small particles, who admits at Paragraph 7 that observational epidemiological studies can't prove causation. I would emphasize that Dr. Devlin fails to point out that small association/low Relative Risk or Odds Ratio results don't even achieve the level of association that allows for a researcher to assert a hypothesis of causation. You might say, a robust result from epidemiology isn't proof, a small association result is even less than that.
3. The 25-page Reference Manual chapter on Epidemiology articulates the rules on proof of causation from observational studies and I highlighted those sections on proof of causation, general causation beginning at page 597 et.seq. and specific causation at page 608 et.seq. I will not discuss the scientific deceit that is used so often trying to make statistical significance into a claim that the evidence is reliable—the scientists in the group know the deceit involved in p value cheating—which doesn't make unreliable evidence proof of anything.

4. Two pages of basic information from the website of the GRADE Working Group (<http://www.gradeworkinggroup.org/>), particularly regarding “Grading the quality of evidence and the strength of recommendations”. This international group is focused on the need for reliable epidemiological evidence. In the 9th paper on quality of evidence in epidemiology, on page 2 item 2, read what they say about small associations and why quality of evidence depends on Relative Risks of more than 2. Every researcher in the air pollution business would be out of business if they followed the rules suggested about the strength of association to prove lethality of pollutants.
5. The 14-page September 28, 2012 American Statistical Association Proceedings paper by Dr. James E. Enstrom "Particulate Matter is Not Killing Californians", which he presented on August 1, 2012 to the ASA 2012 Joint Statistical Meeting Section on Risk Analysis. Dr. Enstrom’s analysis of all sources of evidence on PM2.5 deaths relevant to California provides proof that there is no death effect in California. Tables on pages 2331 and 2332 show small associations with confidence intervals that include 1.0. On page 2333, a US map of PM2.5 mortality risk from the 2000 HEI Reanalysis Report of Krewski also shows that there is no small particle death effect in California. This paper is permanently posted on Dr. Enstrom’s website (<http://www.scientificintegrityinstitute.org/ASAS092812.pdf>).
6. My three-page June 8, 2011 letter to CARB about the clownish performance of Michael Jerrett, trying to make the silk purse out of his air pollution research in California, after initially he admitted in public showed no death effect at all. His trickster attitude shows what lots of research money and time can do to put lipstick on a research pig.
7. The 5-page August 13, 2015 SCIENCE manuscript with nine accomplished coauthors, including me, “Particulate Matter Does Not Cause Premature Deaths”, that includes me as a coauthor, provides evidence, wide and deep that the claims of the South Coast researchers are faulty, and unreliable—that there is no death effect to be shown. This manuscript is permanently posted on the National Association of Scholars website (https://www.nas.org/articles/nas_letter).

Conclusions

I have provided a short version of my objections to research used to support the SCAQMD Draft 2016 AQMP claims about PM2.5 premature deaths.

I hope you read the objections see that the Thurston research cannot be cobbled together with the rest of the research, including the flawed conurbation paper of Michael Jerrett in support of any new small particle regulations BECAUSE the research shows that new air regs will not save lives because there are no deaths.

Thurston, Jerrett, and all the papers on air pollution death studies in California show an overall small particle air pollution death effect of ZERO. What you gonna do—change the rules on how to study toxicity to justify more aggressive and burdensome air regs for Southern California—to achieve what? and at what cost?

I will provide the South Coast People with negative responses on their proposed small particle proposals, when necessary, and depending on what you do with the sorry Thurston results.

You also have a big problem with the show horses you have in the air pollution research community, all of them generate small association studies that don't prove up the South Coast claims about deaths. Not at all. In fact the studies show no death effect is likely, or the associations on the studies would be more consistently robust.

The scientists reading know what I am talking about, a pile of studies with no proof of causation at all, not even a whiff of good evidence for arguments about deaths makes a good argument that the portfolio is my evidence that South Coast is making claims that are not supported by good evidence—and then must fail the smell test.

Dr. Thurston and his now very old small particles paper that admits extremely small Hazard Risks and even Confidence Intervals that include 1.0 is no proof. The Jerrett conurbation gambit is silliness, expensive silliness, but still no proof of a death effect.

I am happy to expand on this letter and attachments by webinar, teleconference or further correspondence in response to questions.

Please make sure this letter and the attachments are made available to the SCAQMD Governing Board.

Thank you.

John Dale Dunn MD JD
Consultant Emergency Services/Peer Review
Civilian Faculty, Emergency Medicine Residency
Carl R. Darnall Army Med Center
Fort Hood, Texas
Medical Officer, Sheriff Bobby Grubbs
Brown County, Texas
325 784 6697 (h) 642 5073 (c)

John Dale Dunn MD JD

Diplomate ABEM, ABLM

Admitted but inactive, Texas and Louisiana Bars

Civilian Contract Faculty, Emergency Medicine Residency Program

Carl R. Darnall Army Medical Center, Fort Hood, TX

Medical Officer, Brown County Sheriff George Caldwell Jr.

401 Rocky Hill Road

Lake Brownwood, Texas 76801

Phone 325 784-6697

E-mail jddmdjd@web-access.net

January 19, 2016

Henry A. Roman, M.S. Industrial Economics, Incorporated (IEc) har@indecon.com

To: Henry A. Roman <har@indecon.com>

CC: George D. Thurston <George.Thurston@nyumc.org>;

Lisa A. Robinson <robinson@hsph.harvard.edu>; Eric D. Ruder <er@indecon.com>

Re: The Proposed 2016 SCAQMD AQMP relies on deceptive human effects research claims and should be scrapped

Mr. Roman,

I will get to the point. Your supportive documents cite the work of George Thurston and in his paper he admits that he finds no evidence that Small Particle Air pollution is killing anyone. When the confidence interval crosses a relative risk of 1.0 all honest scientists declare a null effect.

George Thurston PhD and Co Authors can't find a small particle effect.

My position is that The September 15, 2015 *EHP* paper by Thurston, et al., found NO relationship between PM_{2.5} and total mortality during 2000-2009 in the publicly available NIH-AARP Diet and Health cohort (<http://ehp.niehs.nih.gov/1509676/>).

In the teased out data sets of the study Dr. Thurston tries, with his co authors, to make a silk purse out of pigs ear, because he found some subset data from carved out groups where the usual (for EPA air pollution epidemiologist could be found. But the pig's ear is still there—his findings are small non proof associations for those subgroups, the usual EPA offal, not proof and an overall result of NO EFFECT.

Here's the important section of the abstract with my comments inserted in bold parens to show why the paper does not support the South Coast project to push more small particle regs:

Results: PM_{2.5} exposure was significantly associated with total mortality (HR= 1.03, 95% CI=1.00, 1.05) (**overall CI includes 1.0—no effect**) and CVD mortality (HR=1.10, 95% CI=1.05, 1.15), but the association with respiratory mortality was not statistically significant (HR=1.05, 95% CI=0.98, 1.13) **Authors misused statistically significant, here because it only means they had a desired p value, not results that proved anything**. A significant (**misused again**) association was found with respiratory mortality only among never smokers (HR=1.27; 95% CI: 1.03, 1.56). Associations with 10 µg/m³ PM_{2.5} exposures in yearly participant residential annual mean, or in metropolitan area-wide mean, were consistent with baseline exposure model results. Associations with PM_{2.5} were similar when adjusted for ozone exposures. Analyses of California residents alone

also yielded statistically significant PM_{2.5} mortality HR's for total and CVD mortality

(Not so, small associations don't prove anything, such as HR of 1.03 and 1.1 and anytime the small association is associated with a CI that includes 1.0, no effect can be asserted. And to repeat, all the findings in this study were statistically significant, the negative findings of no effect and the miniscule findings of a small positive effect—the authors intentionally deceive, but they follow a pattern in all air pollution studies of misusing the concept of statistical significance.)

Conclusions: Long-term exposure to PM_{2.5} air pollution was associated with an increased risk of total **(not true, CI included 1.0, miniscule non proof HR)** and CVD mortality **(again, not true, no proof from a small association, and other problems with parsing out a subset)** providing an independent test of the PM_{2.5} – mortality relationship in a new large U.S. prospective cohort experiencing lower post-2000 PM_{2.5} exposure levels. **(Again, small associations don't prove anything and CI that includes 1.0 is null effect. Not only that, but I would suggest that Thurston and colleagues fail the test when they don't advise that their study**

I also object strongly to the misuse of the words that Thurston and co-authors pick to describe their results “statistically significant,” a term of art intentionally designed to put lipstick on a pig. Statistical significance is used by these EPA air pollution researchers to imply valid—however it is nothing more than a method for preventing randomness errors in data management and has nothing to do with the strength or validity of the results. For example in this case a statistically significant result of HR of 1.03 is no proof of anything in a population study, it is not even good enough to be hypothesis generating and requiring stronger or better evidence. As for a statistically significant result (by the data management test of p values) has a Confidence Interval that includes 1.0, the study is proof of nothing, it is a study with a null effect.

To parse out data to find a positive HR in CVD deaths is a deception too—in desk top death certificate tallies CVD deaths dominate but do not actually reflect a diagnosis, just a very uncertain guess. It does provide an opportunity to find a small association; however, that means nothing about proof of causation.

Thurston, and colleagues, being ingenious and they are working for a regulatory entity, so they sliced and diced the data and found—voila—a way to tease out a small effect, admittedly a non proof small effect, that evaded the doom of a CI that included 1.0. It means nothing and is a trick. Shame on them. They aren't finding anything, they were just reworked the data piles to get to a HR that was enough to avoid the nullifying CI that included 1.0. Nice going, but still pseudo-science, because it requires believing in an HR of 1.1.

Since Dr. Thurston and his colleagues don't really know a mechanism for small particles at ambient levels can kill people, another data phenomenon deserves a comment—the CVD results showed a miniscule effect, but the

Respiratory Deaths showed an overall no effect—BUT there was a data surprise, they found a nonsmoker HR that was positive with a relatively large (CI goes up when sample size goes down) and the CI stayed above 1.0 so they could use the magic words “statistically significant” in their deceptive way.

I will not belabor the obvious point that such a non sequitur deserves interesting and a measure of the uncertainties of population studies why the researchers are digging around in effects measures by HRs that are so small as not to deserve attention.

My conclusion is that Thurston and his co-authors were, no doubt, well paid by the NIH and had nothing to offer for the enviro agenda with their study—they are my exhibit one to prove the South Coast needs to reconsider its air pollution rags and reduce the burden on the residents.

I would also remind the South Coast officials that the Thurston study was a 6 state study that obviously must be considered in view of the California experience that will be outlined below—California, even Southern California where air pollution is higher than many other locations, shows no death effect when one assesses the deaths in California cohorts separately.

Michael Jerrett is one of Dr. Thurston’s co-authors, and I am sure he could wax eloquent on the California null effect, since he has been running away from it for a long time. I also suspect that there is a California cohort that could be extracted pretty easily from the Thurston study (it’s called zip codes) and studied and it would show the same null effect.

Imagine, to finish this section off—imagine the weak study Thurston and the almost dead certainty that the Thurston study would show no, nada, nunc effect in California. What’s your guess, Mr. Roman?

Guess what, Mr. Roman, there is a California cohort in the Thurston Study and it shows—just what I said, no effect

In 160,000 deaths in CA here’s the result provided by the Thurston et.al Table 3.

Full Baseline Model for California Only
160,209 deaths Results HR CI
All deaths--1.02 (0.99, 1.04)
CVD deaths 1.10 (1.05, 1.16)
Respiratory deaths 1.01 (0.93, 1.10)

Again, no proof where the CI doesn’t include 1.0, and two parts of the cohort where CI includes 1.0.

The Thurston Study doesn’t pass the smell, taste or laugh test for proof of ambient air pollution caused deaths.

EPA Research Scientist--Under Oath, Epidemiology can't prove our case—we need human exposures

To support that position and remind you, Mr. Roman, and Dr. Thurston, I provide **Appendix A**, attached to this emailed letter, a statement under oath by a Senior EPA official Robert Devlin PhD on the value of epidemiology studies in proving toxicity of air pollutants. I have highlighted for your convenience parsed out sections of his research where he admits epidemiology cannot prove causation, which is the reason the EPA funded attempts to find toxicity with human exposure experiments.

In his declaration under oath Dr. Devlin explains why he is heading up an EPA sponsored human exposure experiments project:

7. Epidemiological observations are the primary tool in the discovery of risks to public health such as that presented by ambient PM_{2.5}. However, epidemiological studies do not generally provide direct evidence of causation. They indicate the existence or lack of a statistical relationship between ambient levels of PM_{2.5} and adverse health outcomes. Large population studies cannot assess the biological mechanisms (called biological plausibility) that could explain how inhaling ambient air pollution particles can cause illness or death in susceptible individuals. This sometimes leaves open the question of whether the observed association in the epidemiological study is causal or whether PM_{2.5} is merely a marker for some other unknown substance.

Controlled human exposure studies offer the opportunity to study small numbers of human subjects under carefully controlled exposure conditions and gain valuable insights

Case 1:12-cv-01066-AJT-TCB Document 14-1 Filed 10/04/12 Page 15 of 135
PageID# 325

into both the relative deposition of inhaled particles and the resulting health effects. Individuals studied can range from healthy people to individuals with cardiac or respiratory diseases of varying degrees of severity. In all cases, the specific protocols defining the subjects, the exposure conditions, and the evaluation procedures must be reviewed and approved by institutional review boards providing oversight for human experimentation. The exposure atmospheres studied vary, ranging from well-defined, single-component aerosols (such as black carbon or sulfuric acid) to atmospheres produced by recently developed particle concentrators, which concentrate the particles present in ambient air. The concentrations of particles studied are limited by ethical considerations and by concern for the range of concentrations, from the experimental setting to typical ambient concentration, over which findings need to be extrapolated.

Exhibit 1 at 36. Controlled human exposures studies have been conducted for decades on important pollutants such as ozone, particulate matter, nitrogen dioxide (NO₂), sulfur dioxide (SO₂), VOCs emitted in from new homes, and carbon monoxide (CO).

9. Controlled human exposure studies assess the biological plausibility of the associations observed in the large-population epidemiological studies. Controlled human exposure studies usually compare the response of an individual following exposure to clean air with their response following exposure to a pollutant that was generated or prepared under carefully controlled conditions, thus providing direct

causal evidence that observed effects are related to the pollutant of interest. These studies are done under conditions that are controlled to ensure safety, with measurable, reversible physiological responses. They are not meant to cause clinically significant adverse health effects, but rather reversible physiological responses can be indicators of the potential for more serious outcomes in susceptible populations identified in epidemiology studies.

I would comment that the human exposure experiments were and are sponsored and funded by EPA in spite of the testimony by EPA officials and before the US Congress as well as public pronouncements by EPA. By the EPA that small particles are lethal, at any level of exposure that would make the exposure experiments illegal, unethical and prohibited by federal statute and American common law as well as international accords on human experiments.

Reference Manual on Scientific Evidence explains the rules on epidemiology.

As a reminder of the rules of epidemiology that Dr. Thurston and your group as well as South Coast official should know, I attach as **Appendix B** the Chapter on Epidemiology in the *Reference Manual on Scientific Evidence* published by the National Academy of Science press and supervised by the Federal Judicial Center. The pertinent parts of the chapter on strength of association are highlighted.

As examples of the points made, from page 602:

*B. How Strong Is the Association Between the Exposure and Disease?*¹⁵⁵

The relative risk is one of the cornerstones for causal inferences.¹⁵⁶ Relative risk measures the strength of the association. The higher the relative risk, the greater the likelihood that the relationship is causal.¹⁵⁷ For cigarette smoking, for example, the estimated relative risk for lung cancer is very high, about 10.¹⁵⁸ That is, the risk of lung cancer in smokers is approximately 10 times the risk in nonsmokers.

A relative risk of 10, as seen with smoking and lung cancer, is so high that it is extremely difficult to imagine any bias or confounding factor that might account for it. The higher the relative risk, the stronger the association and the lower the chance that the effect is spurious. Although lower relative risks can reflect causality, the epidemiologist will scrutinize such associations more closely because there is a greater chance that they are the result of uncontrolled confounding or biases.

And from page 612:

Some courts have reasoned that when epidemiologic studies find that exposure to the agent causes an incidence in the exposed group that is more than twice the incidence in the unexposed group (i.e., a relative risk greater than 2.0), the probability that exposure to the agent caused a similarly situated individual's disease is greater than 50%.¹⁹¹ These courts, accordingly, hold that when there is group-based evidence finding that exposure to an agent causes an incidence of disease in the exposed group that is more than twice the incidence in the unexposed group, the evidence is sufficient to satisfy the plaintiff's burden of production and permit submission of specific causation to a jury. In such a case, the factfinder may find that it is more likely than not that the substance caused the particular plaintiff's disease. Courts, thus, have permitted expert witnesses to testify to specific causation based on the logic of the effect of a doubling of the risk.¹⁹²

GRADE Working Group work on strength of evidence.

I also attach a paper by the highly regarded international public/private scientific group studying integrity in medical research science, called the GRADE Working Group (**Appendix C**), and the paper the discusses their guidelines for strength of evidence, with specifics on how to grade evidence for reliability. In the paper 9 of the series they produced they go to those specifics and I would recommend the paper for your review, Mr. Roman and the review of Dr. Thurston. The GRADE Guidance specifies in its quality of evidence discussion the importance of Relative Risk of 2 or more and the more the better. For proof of benefit the guidance is for a RR of 0.5 or less.

At item 2 on page 2 of the 9th paper in a series of articles produced by the GRADE Working Group for the *Journal of Clinical Epidemiology* (**Appendix C1**) the Authors detail the importance of robust Relative Risk, above 2.0 or below 0.5 as they outline in an adjacent table:

Table 1. Factors that may increase the quality of evidence

1. Large magnitude of effect (direct evidence, relative risk [RR] 2.0 to 5.0 or RR 0.5 < with no plausible confounders); very large with RR 2 to 5 or RR 0.5 or less and no serious problems with risk of bias or precision (sufficiently narrow confidence intervals); more likely to rate up if effect rapid and out of keeping with prior trajectory; usually supported by indirect evidence.
2. Dose-response gradient.
3. All plausible residual confounders or biases would reduce a demonstrated effect, or suggest a spurious effect when results show no effect.

Human experiments by EPA sponsored researchers have not been shown to support their claims that small particles kill—nor have EPA researchers been able to kill animals with extraordinary small particle air pollution exposures.

I not only assert that Dr. Thurston's study shows no evidence to prove deaths by small particles, but I would assert that all the portfolio of EPA sponsored studies on small particles fail to prove deaths because of the same flaws—small associations that prove nothing, no bench science to even suggest a mechanism of death and severely dishonest data torturing that I will explain hereunder.

The flawed EPA research portfolio on human effects of small particles.

There is a compelling listing of the California specific data on small particles pollution and death in all the major studies that are claimed to be proof of lethality. To find a segment of the population not effects is severely damaging to the EPA and CA EPA regime of regulatory efforts to control small particles.

James Enstrom, epidemiologist whom I have worked with to try to stop the research misconduct outlined above, did an analysis of the California cohorts from all the major the major studies that could be mined to separate out California cohorts. Enstrom found a stunning lack of small particle effect in California as demonstrated in the tables below and the dramatic Krewski map of the US showing a decline in small

particle effects from highs in the Eastern US to lows and no effect in the West, including California, thought to have the worst air pollution in the nation. (**Appendix D**)

Shocking news, if you look at the Enstrom California cohort table below. The table of studies has stunningly negative results with the confidence interval of all but 3 of the studies crossing RR of 1.0. Game over, Mr. Roman. The Krewski Map shows no effects in California.

I suggest you Mr. Roman, and Dr. Thurston and his coauthors review this paper that has the null effect information, presented by Dr. Enstrom September 28, 2012 American Statistical Association 2012 JSM Proceedings Session Description and Enstrom Paper on "PM Not Killing CA" (<http://www.scientificintegrityinstitute.org/ASAS092812.pdf>) à find table on PM2.5 and total deaths in CA

For your convenience, Mr. Roman, I have inserted the link for Appendix D that shows the table of California cohorts from the EPAs favorite small particle air pollution studies where California cohorts could be separated. The California data pull was analyzed for RR and Confidence intervals by Dr. Enstrom and it shows a stunning pattern of NULL EFFECT of small particles on deaths.

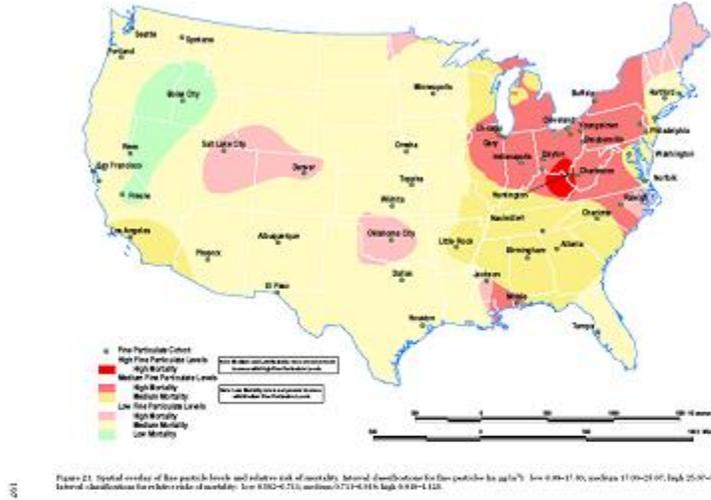
See the tables on the next two pages and the Krewski Map on the third page. The pages are extracted from the document pages 2331-33.

Table 1. Epidemiologic Cohort Studies of PM2.5 and Total Mortality in California
<http://www.scientificintegrityinstitute.org/Enstrom081512.pdf>
 Relative risk of death from all causes (RR and 95% CI) associated with increase of
 10 $\mu\text{g}/\text{m}^3$ in PM2.5

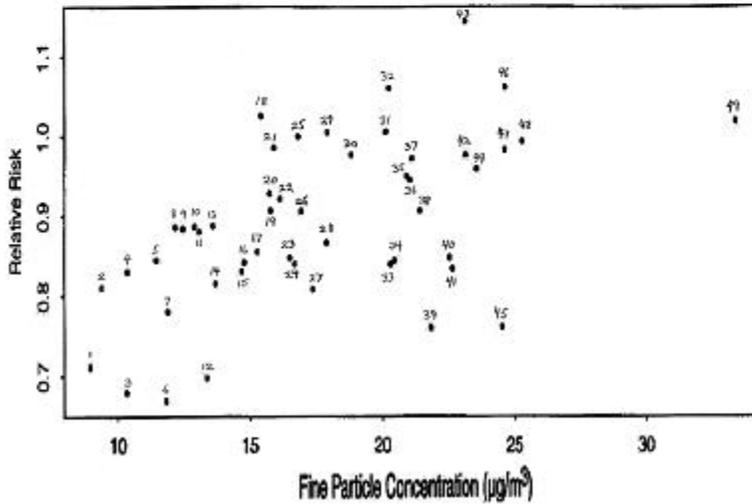
Krewski 2000 & 2010	CA CPS II Cohort (N=40,408 [18,000 M + 22,408 F]; 4 MSAs; 1979-1983 PM2.5; 44 covariates)	RR = 0.872 (0.805-0.944)	1982-1989
McDonnell 2000	CA AHSMOG Cohort (N=3,800 [1,347 M + 2,422 F]; SC&SD&SF AB; M RR=1.09(0.98-1.21) & F RR=0.98(0.92-1.03))	RR ~ 1.00 (0.95 – 1.05)	1977-1992
Jerrett 2005	CPS II Cohort in Los Angeles Basin (N=22,905; 267 zip code areas; 1999-2000 PM2.5; 44 cov + max confounders)	RR = 1.11 (0.99 - 1.25)	1982-2000
Enstrom 2005	CA CPS I Cohort (N=35,783 [15,573 M + 20,210 F]; 11 counties; 1979-1983 PM2.5; 25 county internal comparison)	RR = 1.039 (1.010-1.069) RR = 0.997 (0.978-1.016)	1973-1982 1983-2002
Enstrom 2006	CA CPS I Cohort (N=35,783 [15,573 M + 20,210 F]; 11 counties; 1979-1983 & 1999-2001 PM2.5)	RR = 1.061 (1.017-1.106) RR = 0.995 (0.968-1.024)	1973-1982 1983-2002
Zeger 2008	MCAPS Cohort “West” (3.1 M [1.5 M M + 1.6 M F]; Medicare enrollees in CA+OR+WA (CA=73%); 2000-2005 PM2.5)	RR = 0.989 (0.970-1.008)	2000-2005
Jerrett 2010	CA CPS II Cohort (N=77,767 [34,367 M + 43,400 F]; 54 counties; 2000 PM2.5; KRG ZIP; 20 ind cov+7 eco var; Slide 12)	RR ~ 0.994 (0.965-1.025)	1982-2000
Krewski 2010	CA CPS II Cohort (N=40,408; 4 MSAs; 1979-1983 PM2.5; 44 cov) (N=50,930; 7 MSAs; 1999-2000 PM2.5; 44 cov)	RR = 0.960 (0.920-1.002) RR = 0.968 (0.916-1.022)	1982-2000 1982-2000
Jerrett 2011	CA CPS II Cohort (N=73,609 [32,509 M + 41,100 F]; 54 counties; 2000 PM2.5; KRG ZIP Model; 20 ind cov+7 eco var; Table 28)	RR = 0.994 (0.965-1.024)	1982-2000
Jerrett 2011	CA CPS II Cohort (N=73,609 [32,509 M + 41,100 F]; 54 counties; 2000 PM2.5; Nine Model Ave; 20 ic+7 ev; Fig 22 & Tab 27-32)	RR = 1.002 (0.992-1.012)	1982-2000
Lipsett 2011	CA Teachers Cohort (N=73,489 [73,489 F]; 2000-2005 PM2.5)	RR = 1.01 (0.95 – 1.09)	2000-2005
Ostro 2011	CA Teachers Cohort (N=43,220 [43,220 F]; 2002-2007 PM2.5)	RR = 1.06 (0.96 – 1.16)	2002-2007
replaced Ostro 2010	Incorrect 2010 Result:	RR = 1.84 (1.66 – 2.05)	2002-2007

Figure 1. Figures 21 and 5 from HEI Reanalysis Report (Krewski 2000)

**Figure 21 Spatial Overlay of PM_{2.5} Level and Mortality Risk by City (page 197)
Fine Particles and Mortality Risk**



**Figure 5 (Upper Right) Relative Risk for PM_{2.5} and Total Mortality by City (page 161)
All Cause Mortality (Excluding Boise City, Idaho)**



Mr. Roman, my request to you is that you reevaluate the Thurston paper and confirm what Dr. Thurston admits is no effect in his overall study. Then you must reject participation in any effort by the staff of South Coast to cobble together a case for more onerous regulations of small particles using the Thurston Paper or the Jerrett research of the last few years. You can see that Jerrett's studies in the tables show no effect.

I have written my comments to CARB on the Jerrett conurbation study and discuss it below. Under no circumstances should South Coast burden the citizens of the region based on the Thurston and Jerrett studies on small particle effects.

Since the death effects are projected by your studies from small particles, I will not address the arguments against accepting the other studies on ozone that are not in your area of activity.

Dr. Jerrett admits no effects at the big show in Sacramento.

Dr. Michael Jerrett, prominent air pollution researcher for CARB and EPA (UC Berkeley) had his head handed to him at the public debate/symposium on small particles, Sacramento CAL EPA offices February 26, 2010 as I narrated and told the tale here:

http://www.americanthinker.com/articles/2010/03/californias_toxic_air_scare_ma.html

The 7 hour symposium/debate on small particles is on the video here:

[http://www.cal-span.org/media.php?folder\[\]=CARB](http://www.cal-span.org/media.php?folder[]=CARB)

At the debate, Dr. Jerrett admitted that he couldn't find a death effect in his studies of recent years. He admitted he could not show a death effect and he and CARB hired experts lost the debate to an expert group including Dr. Enstrom, UCLA, Roger McClellan, former Chair of the US EPA Clean Air Scientific Advisory Committee (CASAC), Suresh Moolgavkar MD, U of Washington Cancer Center.

A year and 3 quarters of a million dollars of CARB money later, Dr. Jerrett delivered what he couldn't deliver at Sacramento by what I would describe as flagrant pseudo-science that now is becoming very stylish in junk science for government circles—he did computer models till he could find one that gave him what he wanted. He added, along the way, a prominent list of co-authors:

Principal Investigator:

Michael Jerrett, PhD

Co-Investigators:

Richard T. Burnett, PhD, Arden Pope III, PhD, Daniel Krewski, PhD

George Thurston, ScD, George Christakos, PhD, ScD

Edward Hughes, PhD, Zev Ross, MS, Yuanli Shi, MD, Michael Thun, MD

Funny thing is he admitted his methodology which makes him some kind of evolution of stupid by him and all these prominent air pollution researchers. In his paper reporting small particle effects he and his many prominent co-authors reported positive effects from a parsing of the population data on a temporal spatial template called "conurbation" that showed a positive death effect from small particle air pollution. He also reported the other models he used did not show any effect, and there were 8. So one modeling template gave the group what they wanted and they canned the other 8:

Such scientific strategies are risible, since repeating and confirming is the normal process, but I admit it

gave me a big fat target with a group of charlatans, and I took advantage of the opportunity. See **(Appendix E)**.

Even with data mining and torturing, Dr. Jerrett found an association that was so small it was not proof of causation at all—just like the studies above with RR and HR of less than 1.1.

I had an easy time of it, making fun of Dr. Jerrett's and his high powered group's scientific misconduct, since he and his paper were his own worst enemies. Dr. Jerrett admitted he used multiple models to torture the data until it yielded his desired result, a small, I repeat small, association that would not, in proper application of the rules on magnitude of Relative Risk for epidemiological studies, outlined above, be considered proof of toxicity or lethality, or anything at all.

No matter, the CARB and US EPA and all the anxious advocates of efforts to reduce human activity would believe anything that Dr. Thurston or Dr. Jerrett claimed, even claims of thousands, even hundreds of thousands of lives saved from premature death.

Here are a few excerpts from my letter criticizing the conurbation paper—and I stand by those criticisms here:

My goodness, the subornation gambit is just another form of the well-known researcher trick of chopping the data under multiple methodologies until one finds the result desired with the computer, the mindless computer rigged to find that good result. Changing the geographic parameters to an urban and suburban mix to get a desired effect is bad science that produces outcome based junk.

The rules haven't changed. Dr. Jerrett can't tell us why or how small particles cause disease, so he's short on plausibility; he's also short on specificity because he just uses crude deaths in excess of the predicted and calls them premature. He also, even with such loose methodology, can only show effects in the range under 1.2, so he doesn't have an adequate magnitude of effect to claim proof of causation. Just because Dr. Jerrett is committed to eliminating pollution of any kind, doesn't mean he can claim he is eliminating a toxin, particularly when one considers the following.

1. The researchers have not even bothered to define the nature of the toxin satisfactorily—small particles is a size, 2.5 microns, but it could be weaponized anthrax or agricultural dust—would anyone claim the two are equally toxic?
2. The researchers do not have exposure information—they also use air pollution monitor information for outside air when people live indoors 90 % of the time and they just average it and use it as an exposure index—when will such nonsense be stopped?
3. The decision to use crude death rates and arbitrary short lag times for endpoint of “premature” deaths ignores the nature of chronic diseases. Low level air pollution does not acutely poison people. People die after long periods of illness or disease and failed medical treatment, not some acute exposure to a few microns in a cubic meter of air. What are the researchers studying, is it a real disease or toxic effect or just variable death rates in a population?
4. Premature deaths from what disease, what toxic effect? Specificity is a surrogate in toxicology for plausibility, but it is a separate, important consideration—how can Dr. Jerrett just use premature deaths as an endpoint when we have yet no biologically or toxicologically plausible mechanism for deaths from ambient levels

of air pollution. Dr. Jerrett could be counting deaths from any one of a number of confounding causes.

5. If premature deaths are to be the endpoint rather than tissue proven or test proven disease, when will Dr. Jerrett and his colleagues admit to the problem that they torture crude death rate data for short term rate increases that might correlate with air pollution increases? What proof is that? If they are wrong, a pile of studies that result from such data torturing to find associations is just another extraordinary example of a pattern of research where the principles can't differentiate the noise (death rate variability) from the signal (whatever deaths that might be attributable to air pollution). Monitor information in the range of the noise created by variability of the death rates, lack of real exposure and toxicity information, and arbitrary lag times provide great opportunities for trolling through the data for a correlation.

Could it be that Dr. Jerrett was trolling with the good ship conurbation?

6. If death rates vary as much as 15 percent in populations from winter to summer and variability of death rates from day to day can easily be that much, is Dr. Jerrett, *sans* biological plausibility just reporting on the noise and claiming it is a signal. If the results are in the low range, how much noise, how much signal?

7. If the effect reported fails to meet the *Reference Manual recommendation* that effects be at least 100 percent to be adequate for proof of toxicity, is the Jerrett study just another hypothesis generating study under the rules or another supportive study for the needs of the agency and the air pollution regulatory agenda?

8. Is this conurbation model anything more than a sophisticated form of confirmation bias driven by intellectual passion and commitment with tunnel vision?

9. Is Dr. Jerrett falling for the well-established problem in the air pollution human health effects science community of intellectual passion and commitment combined with confirmation bias and the faggot fallacy? (That faggot fallacy is discussed in *Judging Science* by Huber and Foster (MIT press 1997), and it is the fallacy based on the "belief that multiple pieces of evidence, each independently being suspect or weak, provide strong evidence when bundled together.")

10. Given the source of funding and the CARB commitment to regulating small particles, does anyone on the review panel think Dr. Jerrett would ever, ever receive funding from US EPA or CARB if he repeated his candid admission of February 26, 2010 that would shut down the CARB particle control industry and shut down the CARB and US EPA juggernaut?

And

Cargo Cult Science in the Movie Capital State

I would ask that the reader consider the old and amusing story of Cargo Cults—the mistaken notion of primitives that if they followed some of the appearances of old air fields in South East Asia after the war was over, the planes would return with the people who flew them. Cargo cult science is a fallacious conduct, the pretentious display of scientific customs and methodology that has no substance and is unreliable and unscientific.

The many PhDs arrayed in this very expensive study, even if they presented themselves solemnly and wore white coats, would be involved in a data dredging charade. Bad science cannot be hidden like a Potemkin village, because in the end it's still about the reliability and the credibility of the evidence. Dr. Jerrett's evidence is the great example of the old Texas saying often wrong but never in doubt.

I won't belabor the history and the previous studies that will be brought to the reader's attention about California studies that show no effect. Use of the word significantly might be over the top.

1. A major study by the Health Effects Institute shows no excess mortality from fine particles.
2. The Enstrom Study of a robust cohort of Californians studied over a significant period of time shows no death effect from small particles.
3. The US EPA 2002 report of diesel exhaust health effects showed no effect.
4. The previously mentioned Pope second half data and the Krewski map of effects shows that California residents are not suffering any adverse effects from air pollution.

A good honest study that disproves a hypothesis is controlling—it is evidence that the premise is wrong. Consensus science, a vote of the paid researchers present, or a reliance on authority offends the rules of science—a process that must first of all hold skepticism rather than acquiescence in high regard. Unfortunately hundreds of thousands of dollars from agency coffers can influence research and eliminate self-examination, skepticism and most of all humility and adherence to the rules of science even when it goes against one's personal interests. 8

Scientists must be committed to a careful and skeptical search for truth and reliable results and solutions; they can't become tools of political interests.
Hello—any scientists on watch at CARB or CA EPA?

Economics analysis

I will not spend much time in this letter discussion the inappropriate economics risk/benefits conclusions that come from creating out of whole cloth deaths that never happened and attaching them to a value of almost 10 million per, all to prove up the value of controlling small particles. Mr. Roman, you and I both know that the benefits side of the balance sheet becomes insignificant and not enough to support burdensome regulations if the economists can't put a multimillion dollar value on the specious and unsupported claims of thousands of deaths in the South Coast catchment population.

Let us agree that if you can't prove that the research shows deaths, the economics analyses are worthless exaggerated exercises in releasing agit-prop to the accepting and supportive CA media. The claims certainly overestimate claims of injury by orders of magnitude.

Conclusion

Overwhelming evidence shows that the IEc documents misrepresent and exaggerate the relationship of PM2.5 and ozone to total mortality in the South Coast Air Basin (SCAB) and California. I have explained why. Your faulty claims are embedded in:

“IEc Literature Review of Air Pollution-Related Health Endpoints and Concentration-Response Functions for Particulate Matter: Results and Recommendations Draft Report December 4, 2015”

(http://www.aqmd.gov/docs/default-source/Agendas/STMPR-Advisory-Group/december-2015/3a_draft_pm.pdf?sfvrsn=4)

and

the PPT “IEc Review of Health Endpoints and Economic Valuation for Socioeconomic Report on 2016

South Coast AQMP” (http://www.aqmd.gov/docs/default-source/Agendas/STMPR-Advisory-Group/december-2015/3f_stmpr_presentation_121015.pdf?sfvrsn=4).

I have been working on the problem of bad epidemiology used by CARB and its allies for many years now, trying to cultivate better scientific approaches and fewer panicky exercises in bloviating by EPA advocacy mavens. I would be happy if they would just stop publishing papers that don’t prove anything.

I have been in the practice of medicine, mostly emergency medicine, for just short of 44 years now and I am yet to witness a death from small particles—how bout that? Dr. Thurston is a desk bound person, and I own a stethoscope—he counts death certificates and I fill them out when asked and I assure you that the autopsy rate and the methods displayed by EPA researchers make for epidemiology that really isn’t reliable science.

The rules are still in my favor as explained above. My submission to CARB in the 2008 battle over small particle regulations covers the same ground as this letter. The scientific misconduct of EPA CARB and South Coast sponsored researchers is the same now as many years ago. They violate basic rules of epidemiology to create unreal and unreliable, exaggerated claims of deaths to panic the people and intimidate the policy makers and politicians.

Since I have never seen a person die from American ambient air pollution I condemn and disapprove the the death certificate desk exercises of the EPA, CARB and South Coast researchers as sham science , not real investigations of causes of death. The studies are soaked in deceitful methods and data torturing that result in false assertions and scare mongering for political advantage and to promote an aggressive policy agenda that harms the citizens. **Nothing has changed in 20 years, just more deceptions and more junk science epidemiology paid for by CARB, South Coast, CA EPA and US EPA. The Rules of epidemiology haven’t changed, just the number of times the rules were broken by researchers funded by the EPA.**

In 2012 James Enstrom and other doctoral-level scientists submitted detailed public comments to SCAQMD, which are shown on pages 213-254 of the 2012 AQMP Appendix I Health Effects Document (<http://www.aqmd.gov/docs/default-source/clean-air-plans/air-quality-management-plans/2012-air-quality-management-plan/final-carb-epa-sip-dec2012/2012-aqmp-carb-epa-sip-submittal-appendix-i.pdf>).

These comments provide overwhelming evidence as of 2012 that there is NO relationship between PM2.5 and total mortality in California.

James Enstrom put together an in depth study of the issue in a submission to Science that I support and agree with. The evidence that “Particulate Matter Does Not Cause Premature Deaths” is now even stronger, as summarized in my August 17, 2015 submission to *Science* (<https://www.nas.org/images/documents/PM2.5.pdf>).

I assure you that you, your associates and Dr. Thurston are well advised to inform South Coast officials of my letter and my assertions—more importantly Dr. Thurston and your group, Mr. Roman, have to be honest and forthcoming—you should inform the South Coast Board about the weakness of small associations in epidemiological studies, the lack of bench science to support the claims of small particle lethality, and the null effects of studies on California populations that are found in a focused analysis of the many famous studies of small particle air pollution effects that are referenced above in this letter.

You should also tell the South Coast Board that the studies are piling up to indicate that CA residents don’t suffer from any effects of small particle pollution.

I would also advise you to advise South Coast officials not to try to make the Jerrett study a study that justifies the imposition of more regulations that will be a burden on the economy and welfare of the South Coast Citizens.

I hope this letter alerts you to the dangers of deceit in public policy matters and how bad science cannot justify excessive government regulatory regimes. I have previously warned CARB and CA EPA officials that the False Claims Act provides for severe penalties for those who use taxpayer money, received, for example as a grant, to perpetrate a fraud. Treble damages get your attention?

Thank you for your consideration of this letter and I will copy Dr. Thurston and the individuals listed above. I will not be contacting South Coast Officials and Board Members, anticipating your response to indicate you will be forthcoming and honest in your upcoming presentations to South Coast.

Do you promise to be honest, Mr. Roman, or will you continue this charade of bad science I pursuit of panicking the public and promoting more environmental power grabbing? Your choice.

Remember what I said above about treble damages from the False Claims Act—Dr. Thurston’s study was funded by the NIH, which is funded by taxpayers like me.

Cordially,
s/JDunn MD/
John Dale Dunn MD JD

Attached documents

Appendix A, Dr. Robert Devlin admission under oath (10-04-12)
Appendix B, Chapter on Epidemiology *Reference Manual on Scientific Evidence* (2011)
Appendix C GRADE Working Group website information (08-12-16)
Appendix C 1 GRADE Working Group paper 9 of a series
Appendix D Enstrom paper with tables and US map on human death effects from PM2.5 (09-28-12)
Appendix E Letter by Dunn criticizing Jerrett’s conurbation study of CA PM2.5 deaths (06-08-11)

**Appendix A, Dr. Robert Devlin admission under oath
Case 1:12-cv-01066-AJT-TCB Filed 10/04/12
(53 pages)**

DECLARATION OF ROBERT DEVLIN

I, Robert B. Devlin, pursuant to 28 U.S.C. § 1746, declare, under penalty of perjury, that the following statements are true and correct based upon my personal knowledge, experience or upon information provided to me by persons under my supervision:

1. I am a **Senior Scientist (ST) for the Environmental Public Health Division (EPHD), National Health and Environmental Research Laboratory (NHEERL), Office of Research and Development (ORD), U.S. Environmental Protection Agency.** As one of three STs in NHEERL I am expected to be a scientific leader in the area of air pollution research, to define important areas of research, assemble teams to carry out that research and ensure it is completed in a timely manner and published in peer-reviewed journals. I am currently on detail as Acting Associated Director for Health for NHEERL. Prior to my current position, I was Chief of the Clinical Research Branch (CRB) of the EPHD from 1994 – 2008. The CRB is responsible for doing nearly all controlled human exposure studies within NHEERL. I am also acting Director of EPHD (then call Human Studies Division) in 2007; the Director oversees all research in the Division including epidemiology, clinical and in vitro studies. **I was acting National Program**

Director for ORD's Air Research Program in 2000. This position is the lead for developing research plans related to air pollution for all of ORD and representing the program to groups outside the EPA. I hold adjunct faculty appointments at the University of North Carolina (Chapel Hill) and North Carolina State University. I have been engaged in performing controlled human exposure studies as an EPA investigator since 1986. I have authored or co-authored more than 190 scientific articles, 53 of which involved controlled exposure of human volunteers to air pollutants. The quality of my work at EPA has been recognized by several awards, including one gold and 9 bronze medals, and 8 EPA Scientific and Technological Achievement Awards. I have been invited to present my research at more than 100 Universities, Workshops, and International Meetings.

2. I have a B.S. Degree from the University of Texas (El Paso) that was granted in 1969 and a **Ph.D. degree from the University of Virginia** that was granted in 1976. I was a member of the faculty at Emory University (Atlanta) from 1979 – 1986.

3. I have reviewed the Complaint and exhibits filed in the above-captioned case

4. The term particulate matter (PM) covers a broad class of discrete, but chemically and physically diverse, particles that are ubiquitously present in the ambient air and are emitted from different sources such as power plants, mobile sources, biomass burning, and dust generated by mechanical processes. There are three generally recognized modes of PM defined by particle diameter: very small so-called ultrafine particles that result from the primary emissions related to engine combustion and which are usually in close proximity to those sources; large (coarse) particles primarily generated by abrasive processes and from wind-blown dust; and so-called fine particles which derive from combustion by-products that volatilize and quickly condense or from gases (such as sulfur oxides and nitrogen oxides) that react and transform in the atmosphere after

being emitted. PM_{2.5} is roughly synonymous with fine PM, and generally includes all particulate matter with an aerodynamic diameter of 2.5 micrometers or less. 40 CFR § 50.7(a). Principal sources of PM_{2.5} are fossil fuel combustion, including motor vehicle and power plant emissions, natural and anthropogenic biomass burning, as well as other industrial processes such as smelting. The EPA has specific regulations to control levels of both fine and coarse particles.

5. **In December 2009 EPA issued the Integrated Science Assessment (ISA) for Particulate Matter, pursuant to section 108 of the Clean Air Act (CAA), 42 U.S.C. § 7408.** The ISA is an update of prior science assessments of PM, and reflects the state of the science at that time. The ISA was developed after lengthy review by the Clean Air Scientific Advisory Committee, a federally mandated body charged with advising EPA about scientific matters relating to particulate matter and other forms of air pollution. CAA § 109(d)(2), 42 U.S.C. § 7409(d)(2). Development of an ISA typically involves the consideration of thousands of scientific studies conducted in the U.S. and around the world as part of assessing the relationship between air pollutant exposures and health effects. In the ISA, the entire body of scientific evidence, including epidemiological, controlled human exposure, animal toxicological studies, studies with cultured cells, as well as other sources of information, is assessed and an overall judgment is made on the causal relationship between exposure to ambient PM_{2.5} and health effects. The ISA provides the scientific basis for development of the National Ambient Air Quality Standards (NAAQS) for an air pollutant. CAA § 109(b).¹

6. Epidemiological studies typically use data from large populations of people with varying susceptibility to PM_{2.5} and evaluate the relationship between short or long-term changes in ambient levels of PM_{2.5}, e.g. changes in the 24-hour average level of PM_{2.5} measured at

¹ Ambient air refers to outdoor air in places that members of the public have access to. 40 C.F.R. § 50.1.

monitors in a metropolitan area, with changes in mortality and morbidity such as the numbers of emergency department visits and hospital admissions. This generally involves the use of complex statistical methods to evaluate the mathematical relationship between variations in measured ambient air pollution levels and health data.

7. Epidemiological observations are the primary tool in the discovery of risks to public health such as that presented by ambient PM_{2.5}. However, epidemiological studies do not generally provide direct evidence of causation. They indicate the existence or lack of a statistical relationship between ambient levels of PM_{2.5} and adverse health outcomes. Large population studies cannot assess the biological mechanisms (called biological plausibility) that could explain how inhaling ambient air pollution particles can cause illness or death in susceptible individuals. This sometimes leaves open the question of whether the observed association in the epidemiological study is causal or whether PM_{2.5} is merely a marker for some other unknown substance.

8. Controlled human exposure studies conducted by EPA scientists and EPA funded scientists at multiple universities in the United States fill an information gap that cannot be filled by large population studies. In 1998 the Committee on Research Priorities for Airborne Particulate Matter was established by the National Research Council in response to a request from Congress. The committee was charged with producing four reports over a five-year period which describe a conceptual framework for an integrated national program of particulate-matter research and identified the most critical research needs linked to key policy-related scientific uncertainties. Excerpts from their most recent report (published in 2004) are attached as Exhibit

1 to this Declaration. On page 36 the Committee says:

Controlled human exposure studies offer the opportunity to study small numbers of human subjects under carefully controlled exposure conditions and gain valuable insights

into both the relative deposition of inhaled particles and the resulting health effects. Individuals studied can range from healthy people to individuals with cardiac or respiratory diseases of varying degrees of severity. In all cases, the specific protocols defining the subjects, the exposure conditions, and the evaluation procedures must be reviewed and approved by institutional review boards providing oversight for human experimentation. The exposure atmospheres studied vary, ranging from well-defined, single-component aerosols (such as black carbon or sulfuric acid) to atmospheres produced by recently developed particle concentrators, which concentrate the particles present in ambient air. The concentrations of particles studied are limited by ethical considerations and by concern for the range of concentrations, from the experimental setting to typical ambient concentration, over which findings need to be extrapolated.

Exhibit 1 at 36. Controlled human exposures studies have been conducted for decades on important pollutants such as ozone, particulate matter, nitrogen dioxide (NO₂), sulfur dioxide (SO₂), VOCs emitted in from new homes, and carbon monoxide (CO).

9. Controlled human exposure studies assess the biological plausibility of the associations observed in the large-population epidemiological studies. Controlled human exposure studies usually compare the response of an individual following exposure to clean air with their response following exposure to a pollutant that was generated or prepared under carefully controlled conditions, thus providing direct causal evidence that observed effects are related to the pollutant of interest. These studies are done under conditions that are controlled to ensure safety, with measurable, reversible physiological responses. They are not meant to cause clinically significant adverse health effects, but rather reversible physiological responses can be indicators of the potential for more serious outcomes in susceptible populations identified in epidemiology studies. As such, controlled human exposure studies do not study individuals felt to be at significant risk; they almost always study healthy individuals or people with conditions such as mild asthma. Controlled human exposure studies, together with toxicological studies, provide important insights which can improve our understanding of the potential biological mechanisms or pathways for effects observed in epidemiological studies (e.g., respiratory symptoms or

cardiovascular events, hospital admissions or emergency department visits, or premature death).

10. Obtaining information on the biological impacts of exposure to PM_{2.5} from controlled human exposure studies such as the CAPTAIN study is a very important element in developing an integrated body of scientific knowledge to evaluate the impact on health from exposure to PM

2.5 air pollution. The CAPTAIN study is particularly important in that it addresses an area of PM research where there are still important questions related to fully understanding the role of specific components included in the mixtures of fine particles represented by PM_{2.5} that may be more closely related to the cardiovascular health effects observed in epidemiological studies. PM_{2.5} is a complex mixture derived from several different sources. **There is still uncertainty as to which components or sources of PM_{2.5} are most responsible for causing effects people and if different components or sources cause effects by different biological mechanisms. This type of research can help address existing uncertainties in the PM scientific literature, providing important evidence for informing future PM NAAQS reviews and, in particular, consideration of possible alternative particle indicators and/or standard levels.** In some cases, research in these areas can go beyond aiding standard setting to informing the development of more efficient and effective control strategies.

11. For ethical and safety reasons, controlled human exposure studies to air pollution conducted by NHEERL are initiated only if there is evidence that any effects to the subjects resulting from exposure will be mild, transient, and reversible, and if there is prior data from one or more of the following types of research:

a. Testing in laboratory animals.

b. Observational research involving only naturally occurring human exposures.

c. Human studies involving a closely related air pollutant.

12. Based on the entire body of scientific evidence, including epidemiological, controlled human exposure, and toxicological studies, the ISA for PM drew several important conclusions about the relationship between exposure to PM_{2.5} and health effects. For short-term exposures to PM_{2.5}, the ISA concluded there was a causal relationship between ambient PM and cardiovascular effects. The epidemiologic evidence showed that increases in 24-hour levels of ambient PM_{2.5} was mathematically associated with an increase in hospital admission or emergency room visits, predominantly for ischemic heart disease [IHD] and congestive heart failure [CHF]. See ISA p. 2-9, attached as Exhibit 2 to this Declaration. **There was also evidence from a small number of toxicological and controlled human exposure studies that supported the biological plausibility of this conclusion, although these studies needed to be duplicated and expanded to identify specific PM components and sources which are of most concern. The ISA also concluded there was a causal relationship between ambient PM and mortality. An evaluation of the epidemiological literature indicates consistent positive associations between short-term exposure to PM_{2.5} and all-cause, cardiovascular-, and respiratory-related mortality. ISA p. 2-10, Exhibit 2 to this Declaration. Finally, the ISA concluded that there was a likely casual relationship between ambient PM and respiratory effects. The recent epidemiological studies that have been evaluated report consistent positive associations between short-term exposure to PM_{2.5} and respiratory emergency department visits and hospital admissions for chronic obstructive pulmonary disease (COPD) and respiratory infections. ISA p.2-10, Exhibit 2 to this Declaration. The evidence of serious health effects such as hospital admissions, emergency department visits, and death, all derived from a large body of epidemiological studies.**

13. The risk of serious health effects from exposure to typical levels of PM_{2.5} is largely focused on people with preexisting illnesses, such as elderly people with cardiovascular diseases or COPD. Even for people with preexisting diseases, there is no evidence that all persons are affected the same way or have the same degree of risk.

14. The body of scientific evidence also informs us on what risks there are to an individual that is exposed to PM_{2.5}. For example, **it is clear that PM_{2.5} is not lethal or toxic to all people. The risk of serious health effects is clearly focused on people such as those with pre-existing cardio or respiratory illness. When very large numbers of people are exposed, as occurs in major population centers, the overall risk to the public is large enough to present a serious public health problem in the form of increased mortality and morbidity. It is this serious risk to the overall public health that leads EPA to describe PM as a serious public health problem.**

15. However, the risk to an individual is very different from the overall public health risk associated with exposures of large populations of people to ambient air levels of PM_{2.5}. This is especially true if the individual does not have pre-existing health conditions such as preexisting cardiovascular disease. While it is impossible to say there is no risk to a healthy individual, epidemiology studies provide evidence that the risk to healthy individuals is considered to be very small. Institutional review boards (IRBs) are charged with overseeing the safe and ethical conduct of human studies. IRBs from the University of North Carolina Medical School (which oversee EPA studies done on the campus of the University of North Carolina) as well as those which oversee human studies at several universities throughout the US, in Canada, England, and Sweden have all examined the risk posed to individuals exposed to particulate air pollution and concluded that these studies are safe and ethical to perform.

16. EPA relies on the entire body of scientific evidence to draw judgments about the risk to the

public health from exposure to ambient PM. In settings the NAAQS, EPA exercise it scientific and public health judgment and determines levels that will protect the public health, including groups of people that are more at risk to the air pollutant under consideration, with an adequate margin of safety. In the case of PM_{2.5}, the people most at risk from exposure to ambient PM_{2.5} include those with pre-existing cardiovascular illness or respiratory illness. The current NAAQS is 15.0 ug/m³ annual average, and a 35 ug/m³ 24-hour average. The 24 hour average is met if the 3 year average of the 98th percentile is 35 ug/m³ or below. The 98th percentile means that approximately 6 or 7 days in the year can have higher concentrations than the day used to compare to the 35 ug/m³.²

Dated: October 3, 2012

A handwritten signature in black ink, appearing to read "Robert B. Devlin", is written over a horizontal line.

Robert B. Devlin

² The air quality in Chapel Hill, NC, where the subjects are tested, is well within the levels that attain the current NAAQS.

Devlin Declaration

Exhibit 1

Research Priorities for Airborne Particulate Matter

•IV•

Continuing Research Progress

Committee on Research Priorities for
Airborne Particulate Matter

Board on Environmental Studies and Toxicology

Division on Earth and Life Studies

NATIONAL RESEARCH COUNCIL

OF THE NATIONAL ACADEMIES

THE NATIONAL ACADEMIES PRESS Washington, D.C. www.nap.edu

THE NATIONAL ACADEMIES PRESS 500 Fifth Street, NW Washington, DC 20001

NOTICE: The project that is the subject of this report was approved by the Governing Board of the National Research Council, whose members are drawn from the councils of the National Academy of Sciences, the National Academy of Engineering, and the Institute of Medicine. The members of the committee responsible for the report were chosen for their special competences and with regard for appropriate balance.

This project was supported by Contract No. 68-C-98-003 between the National Academy of Sciences and the U.S. Environmental Protection Agency. Any opinions, findings, conclusions, or recommendations expressed in this publication are those of the author(s) and do not necessarily reflect the view of the organizations or agencies that provided support for this project.

Library of Congress Control Number 2004112036

International Standard Book Number 0-309-09199-3 (Book)
0-309-53170-5 (PDF)

Additional copies of this report are available from: The National Academies Press
500 Fifth Street, NW Box 285
Washington, DC 20055

800-624-6242
202-334-3313 (in the Washington metropolitan area)
<http://www.nap.edu>

Copyright 2004 by the National Academy of Sciences. All rights reserved. Printed in the United States of America

but some of the groups considered most susceptible—persons with advanced chronic heart and lung diseases—are not yet well studied. Additionally, the extent to which findings from any particular location can be generalized is uncertain, and many studies to date have focused primarily on total particle mass, rather than more detailed particle characteristics, such as their chemical composition.

Epidemiological studies take advantage of naturally occurring variation in exposure, across groups or over time, to estimate the effect of PM on one or more health outcome indicators. In an effort to provide evidence relevant to the NAAQS for PM, epidemiologists design studies that have the potential to estimate the effect of PM without contamination (confounding) by the effects of other pollutants. This approach implicitly assumes that inhaled particles have effects on health that are independent of other pollutants, an underlying assumption in having a NAAQS for PM. Alternatively, the effect assigned to PM may reflect the total effect of the air pollution mixture or some other factor that varies with PM, and PM is serving as a surrogate index. **Even with careful design and analysis, there is the possibility of some residual confounding of the effect of PM by other pollutants or other factors. Some epidemiological studies take advantage of historical data on air quality and community health. Other studies use air quality and health data collected prospectively to address specific hypotheses.**

Controlled human exposure studies offer the opportunity to study small numbers of human subjects under carefully controlled exposure conditions and gain valuable insights into both the relative deposition of inhaled particles and the resulting health effects. Individuals studied can range from healthy people to individuals with cardiac or respiratory diseases of varying degrees of severity. In all cases, the specific protocols defining the subjects, the exposure conditions, and the evaluation procedures must be reviewed and approved by institutional review boards providing oversight for human experimentation. The exposure atmospheres studied vary, ranging from well-defined, single-component aerosols (such as black carbon⁵ or sulfuric acid) to atmospheres produced by recently developed particle concentrators, which concentrate the particles present in ambient air. The concentrations of particles studied are limited by ethical considerations and by concern for the range of concentrations, from the experimental setting to typical ambient concentration, over which findings need to be extrapolated.

Toxicological studies with laboratory animals provide the opportunity

⁵"Black carbon" is a general term that is often used interchangeably with "elemental carbon" or "soot."

Devlin Declaration
Exhibit 2

Chapter 2. Integrative Health and Welfare Effects Overview

The subsequent chapters of this ISA will present the most policy-relevant information related to this review of the NAAQS for PM. This chapter integrates the key findings from the disciplines evaluated in this current assessment of the PM scientific literature, which includes the atmospheric sciences, ambient air data analyses, exposure assessment, dosimetry, health studies (e.g., toxicological, controlled human exposure, and epidemiologic), and welfare effects. The EPA framework for causal determinations described in Chapter 1 has been applied to the body of scientific evidence in order to collectively examine the health or welfare effects attributed to PM exposure in a two-step process.

As described in Chapter 1, EPA assesses the results of recent relevant publications, building upon evidence available during the previous NAAQS reviews, to draw conclusions on the causal relationships between relevant pollutant exposures and health or environmental effects. This ISA uses a five-level hierarchy that classifies the weight of evidence for causation:

- ⑩ Causal relationship
- ⑩ Likely to be a causal relationship
- ⑩ Suggestive of a causal relationship
- ⑩ Inadequate to infer a causal relationship
- ⑩ Not likely to be a causal relationship

Beyond judgments regarding causality are questions relevant to quantifying health or environmental risks based on our understanding of the quantitative relationships between pollutant exposures and health or welfare effects. Once a determination is made regarding the causal relationship between the pollutant and outcome category, important questions regarding quantitative relationships include:

- ⑩ What is the concentration-response or dose-response relationship?
- ⑩ Under what exposure conditions (amount deposited, dose or concentration, duration and pattern) are effects observed?
- ⑩ What populations appear to be differentially affected (i.e., more susceptible) to effects?
- ⑩ What elements of the ecosystem (e.g., types, regions, taxonomic groups, populations, functions, etc.) appear to be affected, or are more sensitive to effects?

To address these questions, in the second step of the EPA framework, the entirety of quantitative evidence is evaluated to identify and characterize potential concentration-response relationships. This requires evaluation of levels of pollutant and exposure durations at which effects were observed for exposed populations including potentially susceptible populations.

This chapter summarizes and integrates the newly available scientific evidence that best informs consideration of the policy-relevant questions that frame this assessment, presented in Chapter 1. Section 2.1 discusses the trends in ambient concentrations and sources of PM and provides a brief summary of ambient air quality. Section 2.2 presents the evidence regarding personal exposure to ambient PM in outdoor and indoor microenvironments, and it discusses the

⑩ Note: Hyperlinks to the reference citations throughout this document will take you to the NCEA HERO database (Health and Environmental Research Online) at <http://epa.gov/hero>. HERO is a database of scientific literature used by U.S. EPA in the process of developing science assessments such as the Integrated Science Assessments (ISA) and the Integrated Risk Information System (IRIS).

relationship between ambient PM concentrations and exposure to PM from ambient sources. Section 2.3 integrates the evidence for studies that examine the health effects associated with short- and long-term exposure to PM and discusses important uncertainties identified in the interpretation of the scientific evidence. **Section 2.4 provides a discussion of policy-relevant considerations, such as potentially susceptible populations, lag structure, and the PM concentration-response relationship, and PM sources and constituents linked to health effects. Section 2.5 summarizes the evidence for welfare effects related to PM exposure. Finally, Section 2.6 provides all of the causal determinations reached for each of the health outcomes and PM exposure durations evaluated in this ISA.**

2.1. Concentrations and Sources of Atmospheric PM

2.1.1. Ambient PM Variability and Correlations

Recently, advances in understanding the spatiotemporal distribution of PM mass and its constituents have been made, particularly with regard to PM_{2.5} and its components as well as ultrafine particles (UFPs). Emphasis in this ISA is placed on the period from 2005-2007, incorporating the most recent validated EPA Air Quality System (AQS) data. The AQS is EPA's repository for ambient monitoring data reported by the national, and state and local air monitoring networks. Measurements of PM_{2.5} and PM₁₀ are reported into AQS, while PM_{10-2.5} concentrations are obtained as the difference between PM₁₀ and PM_{2.5} (after converting PM₁₀ concentrations from STP to local conditions; Section 3.5). Note, however, that a majority of U.S. counties were not represented in AQS because their population fell below the regulatory monitoring threshold. Moreover, monitors reporting to AQS were not uniformly distributed across the U.S. or within counties, and conclusions drawn from AQS data may not apply equally to all parts of a geographic region. Furthermore, biases can exist for some PM constituents (and hence total mass) owing to volatilization losses of nitrates and other semi-volatile compounds, and, conversely, to retention of particle-bound water by hygroscopic species. The degree of spatial variability in PM was likely to be region-specific and strongly influenced by local sources and meteorological and topographic conditions.

2.1.1.1. Spatial Variability across the U.S.

AQS data for daily average concentrations of PM_{2.5} for 2005-2007 showed considerable variability across the U.S. (Section 3.5.1.1). Counties with the highest average concentrations of PM_{2.5} (>18 µg/m³) were reported for several counties in the San Joaquin Valley and inland southern California as well as Jefferson County, AL (containing Birmingham) and Allegheny County, PA (containing Pittsburgh). Relatively few regulatory monitoring sites have the appropriate co-located monitors for computing PM_{10-2.5}, resulting in poor geographic coverage on a national scale (Figure 3-10). Although the general understanding of PM differential settling leads to an expectation of greater spatial heterogeneity in the PM_{10-2.5} fraction, deposition of particles as a function of size depends strongly on local meteorological conditions. Better geographic coverage is available for PM₁₀, where the highest reported annual average concentrations (>50 µg/m³) occurred in southern California, southern Arizona and central New Mexico. The size distribution of PM varied substantially by location, with a generally larger fraction of PM₁₀ mass in the PM_{10-2.5} size range in western cities (e.g., Phoenix and Denver) and a larger fraction of PM₁₀ in the PM_{2.5} size range in eastern U.S. cities (e.g., Pittsburgh and Philadelphia). UFPs are not measured as part of AQS or any other routine regulatory network in the U.S. Therefore, limited information is available regarding regional variability in the spatiotemporal distribution of UFPs.

Spatial variability in PM_{2.5} components obtained from the Chemical Speciation Network (CSN) varied considerably by species from 2005-2007 (Figures 3-12 through 3-18). The highest annual average organic carbon (OC) concentrations were observed in the western and southeastern U.S. OC concentrations in the western U.S. peaked in the fall and winter, while OC concentrations in the Southeast peaked anytime between spring and fall. Elemental carbon (EC) exhibited less seasonality than OC and showed lowest seasonal variability in the eastern half of the U.S. The

highest annual average EC concentrations were present in Los Angeles, Pittsburgh, New York, and El Paso. Concentrations of sulfate (SO_4^{2-}) were higher in the eastern U.S. as a result of higher SO_2 emissions in the East compared with the West. There is also considerable seasonal variability with higher SO_4^{2-} concentrations in the summer months when the oxidation of SO_2 proceeds at a faster rate than during the winter. Nitrate (NO_3^-) concentrations were highest in California and during the winter in the Upper Midwest. In general, NO_3^- was higher in the winter across the country, in part as a result of temperature-driven partitioning and volatilization. Exceptions existed in Los Angeles and Riverside, CA, where high NO_3^- concentrations appeared year-round. There is variation in both $\text{PM}_{2.5}$ mass and composition among cities, some of which might be due to regional differences in meteorology, sources, and topography.

2.1.1.2. Spatial Variability on the Urban and Neighborhood Scales

In general, $\text{PM}_{2.5}$ has a longer atmospheric lifetime than $\text{PM}_{10-2.5}$. As a result, $\text{PM}_{2.5}$ is more homogeneously distributed than $\text{PM}_{10-2.5}$, whose concentrations more closely reflect proximity to local sources (Section 3.5.1.2). Because PM_{10} encompasses $\text{PM}_{10-2.5}$ in addition to $\text{PM}_{2.5}$, it also exhibits more spatial heterogeneity than $\text{PM}_{2.5}$. Urban- and neighborhood-scale variability in PM mass and composition was examined by focusing on 15 metropolitan areas, which were chosen based on their geographic distribution and coverage in recent health effects studies. The urban areas selected were Atlanta, Birmingham, Boston, Chicago, Denver, Detroit, Houston, Los Angeles, New York, Philadelphia, Phoenix, Pittsburgh, Riverside, Seattle and St. Louis. Inter-monitor correlation remained higher over long distances for $\text{PM}_{2.5}$ as compared with PM_{10} in these 15 urban areas. To a large extent, greater variation in $\text{PM}_{2.5}$ and PM_{10} concentrations within cities was observed in areas with lower ratios of $\text{PM}_{2.5}$ to PM_{10} . When the data was limited to only sampler pairs with less than 4 km separation (i.e., on a neighborhood scale), inter-sampler correlations remained higher for $\text{PM}_{2.5}$ than for PM_{10} . The average inter-sampler correlation was 0.93 for $\text{PM}_{2.5}$, while it dropped to 0.70 for PM_{10} (Section 3.5.1.3). Insufficient data were available in the 15 metropolitan areas to perform similar analyses for $\text{PM}_{10-2.5}$ using co-located, low volume FRM monitors.

As previously mentioned, UFPs are not measured as part of AQS or any other routine regulatory network in the U.S. Therefore, information about the spatial variability of UFPs is sparse; however, their number concentrations are expected to be highly spatially and temporally variable. This has been shown on the urban scale in studies in which UFP number concentrations drop off quickly with distance from roads compared to accumulation mode particle numbers.

2.1.2. Trends and Temporal Variability

Overall, $\text{PM}_{2.5}$ concentrations decreased from 1999 (the beginning of nationwide monitoring for $\text{PM}_{2.5}$) to 2007 in all ten EPA Regions, with the 3-yr avg of the 98th percentile of 24-h $\text{PM}_{2.5}$ concentrations dropping 10% over this time period. However from 2002-2007, concentrations of $\text{PM}_{2.5}$ were nearly constant with decreases observed in only some EPA Regions (Section 3.5.2.1).

Concentrations of $\text{PM}_{2.5}$ components were only available for 2002-2007 using CSN data and showed little decline over this time period. This trend in $\text{PM}_{2.5}$ components is consistent with trends in $\text{PM}_{2.5}$ mass concentration observed after 2002 (shown in Figures 3-44 through 3-47). Concentrations of PM_{10} also declined from 1988 to 2007 in all ten EPA Regions.

Using hourly PM observations in the 15 metropolitan areas, diel variation showed average hourly peaks that differ by size fraction and region (Section 3.5.2.3). For both $\text{PM}_{2.5}$ and PM_{10} , a morning peak was typically observed starting at approximately 6:00 a.m., corresponding with the start of morning rush hour. There was also an evening concentration peak that was broader than the morning peak and extended into the overnight period, reflecting the concentration increase caused by the usual collapse of the mixing layer after sundown. The magnitude and duration of these peaks varied considerably by metropolitan area investigated.

UFPs were found to exhibit similar two-peaked diel patterns in Los Angeles and the San Joaquin Valley of CA and Rochester, NY as well as in Kawasaki City, Japan, and Copenhagen, Denmark. The morning peak in UFPs likely represents primary source emissions, such as rush-hour traffic, while the afternoon peak likely represents the combination of primary source emissions and nucleation of new particles.

2.1.3. Correlations between Copollutants

Correlations between PM and gaseous copollutants, including SO₂, NO₂, carbon monoxide (CO) and O₃, varied both seasonally and spatially between and within metropolitan areas (Section 3.5.3). On average, PM_{2.5} and PM₁₀ were correlated with each other better than with the gaseous copollutants. Although data are limited for PM_{10-2.5}, the available data suggest a stronger correlation between PM₁₀ and PM_{10-2.5} than between PM_{2.5} and PM_{10-2.5} on a national basis. There was relatively little seasonal variability in the mean correlation between PM in both size fractions and SO₂ and NO₂. CO, however, showed higher correlations with PM_{2.5} and PM₁₀ on average in the winter compared with the other seasons. This seasonality results in part because a larger fraction of PM is primary in origin during the winter. To the extent that this primary component of PM is associated with common combustion sources of NO₂ and CO, then higher correlations with these gaseous copollutants are to be expected. Increased atmospheric stability in colder months also results in higher correlations between primary pollutants (Section 3.5).

The correlation between daily maximum 8-h avg O₃ and 24-h avg PM_{2.5} showed the highest degree of seasonal variability with positive correlations on average in summer (avg = 0.56) and negative correlations on average in the winter (avg = -0.30). During the transition seasons, spring and fall, correlations were mixed but on average were still positive. PM_{2.5} is both primary and secondary in origin, whereas O₃ is only secondary. Photochemical production of O₃ and secondary PM in the planetary boundary layer (PBL) is much slower during the winter than during other seasons. Primary pollutant concentrations (e.g., primary PM_{2.5} components, NO and NO₂) in many urban areas are elevated in winter as the result of heating emissions, cold starts and low mixing heights. O₃ in the PBL during winter is mainly associated with air subsiding from above the boundary layer following the passage of cold fronts, and this subsiding air has much lower PM concentrations than are present in the PBL. Therefore, a negative association between O₃ and PM_{2.5} is frequently observed in the winter. During summer, both O₃ and secondary PM_{2.5} are produced in the PBL and in the lower free troposphere at faster rates compared to winter, and so they tend to be positively correlated.

2.1.4. Measurement Techniques

The federal reference methods (FRMs) for PM_{2.5} and PM₁₀ are based on criteria outlined in the Code of Federal Regulations. They are, however, subject to several limitations that should be kept in mind when using compliance monitoring data for health studies. For example, FRM techniques are subject to the loss of semi-volatile species such as organic compounds and ammonium nitrate (especially in the West). Since FRMs based on gravimetry use 24-h integrated filter samples to collect PM mass, no information is available for variations over shorter averaging times from these instruments. However, methods have been developed to measure real-time PM mass concentrations. Real-time (or continuous and semi-continuous) measurement techniques are also available for PM species, such as particle into liquid sampler (PILS) for multiple ions analysis and aerosol mass spectrometer (AMS) for multiple components analysis (Section 3.4.1). Advances

have also been achieved in PM organic speciation. New 24-h FRMs and Federal Equivalent Methods (FEMs) based on gravimetry and continuous FEMs for PM_{10-2.5} are available. FRMs for PM_{10-2.5} rely on calculating the difference between co-located PM₁₀ and PM_{2.5} measurements while a dichotomous sampler is designated as an FEM.

2.1.5. PM Formation in the Atmosphere and Removal

PM in the atmosphere contains both primary (i.e., emitted directly by sources) and secondary components, which can be anthropogenic or natural in origin. Secondary PM components can be produced by the oxidation of precursor gases such as SO₂ and NO_x to acids followed by neutralization with ammonia (NH₃) and the partial oxidation of organic compounds. In addition to being emitted as primary particles, UFPs are produced by the nucleation of H₂SO₄ vapor, H₂O vapor, and perhaps NH₃ and certain organic compounds. Over most of the earth's surface, nucleation is probably the major mechanism forming new UFPs. New UFP formation has been observed in environments ranging from relatively unpolluted marine and continental environments to polluted

urban areas as an ongoing background process and during nucleation events. However, as noted above, a large percentage of UFPs come from combustion-related sources such as motor vehicles. Developments in the chemistry of formation of secondary organic aerosol (SOA) indicate that oligomers are likely a major component of OC in aerosol samples. Recent observations also suggest that small but significant quantities of SOA are formed from the oxidation of isoprene in addition to the oxidation of terpenes and organic hydrocarbons with six or more carbon atoms. Gasoline engines have been found to emit a mix of nucleation-mode heavy and large polycyclic aromatic hydrocarbons on which unspent fuel and trace metals can condense, while diesel particles are composed of a soot nucleus on which sulfates and hydrocarbons can condense. To the extent that the primary component of organic aerosol is overestimated in emissions from combustion sources, the semi-volatile components are underestimated. This situation results from the lack of capture of evaporated semi-volatile components upon dilution in common emissions tests. As a result, near-traffic sources of precursors to SOA would be underestimated. The oxidation of these precursors results in more oxidized forms of SOA than previously considered, in both near source urban environments and further downwind. Primary organic aerosol can also be further oxidized to forms that have many characteristics in common with oxidized SOA formed from gaseous precursors. Organic peroxides constitute a significant fraction of SOA and represent an important class of reactive oxygen species (ROS) that have high oxidizing potential. More information on sources, emissions and deposition of PM are included in Section 3.3.

Wet and dry deposition are important processes for removing PM and other pollutants from the atmosphere on urban, regional, and global scales. Wet deposition includes incorporation of particles into cloud droplets that fall as rain (rainout) and collisions with falling rain (washout). Other hydrometeors (snow, ice) can also serve the same purpose. Dry deposition involves transfer of particles through gravitational settling and/or by impaction on surfaces by turbulent motions. The effects of deposition of PM on ecosystems and materials are discussed in Section 2.5 and in Chapter 9.

2.1.6. Source Contributions to PM

Results of receptor modeling calculations indicate that $PM_{2.5}$ is produced mainly by combustion of fossil fuel, either by stationary sources or by transportation. A relatively small number of broadly defined source categories, compared to the total number of chemical species that typically are measured in ambient monitoring source receptor studies, account for the majority of the observed PM mass. Some ambiguity is inherent in identifying source categories. For example, quite different mobile sources such as trucks, farm equipment, and locomotives rely on diesel engines and ancillary data is often required to resolve these sources. A compilation of study results shows that secondary SO_4^{2-} (derived mainly from SO_2 emitted by Electricity Generating Units [EGUs]), NO_3^- (from the oxidation of NO_x emitted mainly from transportation sources and EGUs), and primary mobile source categories, constitute most of $PM_{2.5}$ (and PM_{10}) in the East. $PM_{10-2.5}$ is mainly primary in origin, having been emitted as fully formed particles derived from abrasion and crushing processes, soil disturbances, plant and insect fragments, pollens and other microorganisms, desiccation of marine aerosol emitted from bursting bubbles, and hygroscopic fine PM expanding with humidity to coarse mode. Gases such as HNO_3 can also condense directly onto preexisting coarse particles. Suspended primary coarse PM can contain Fe, Si, Al, and base cations from soil, plant and insect fragments, pollen, fungal spores, bacteria, and viruses, as well as fly ash, brake lining particles, debris, and automobile tire fragments. Quoted uncertainties in the source apportionment of constituents in ambient aerosol samples typically range from 10 to 50%. An intercomparison of source apportionment techniques indicated that the same major source categories of $PM_{2.5}$ were consistently identified by several independent groups working with the same data sets. Soil-, sulfate-, residual oil-, and salt-associated mass were most clearly identified by the groups. Other sources with more ambiguous signatures, such as vegetative burning and traffic-related emissions were less consistently identified. Spatial variability in source contributions across urban areas is an important consideration in assessing the likelihood of exposure error in epidemiologic studies relating health outcomes to sources. Concepts similar to those for using ambient concentrations as surrogates for personal exposures apply here. Some source attribution studies for $PM_{2.5}$ indicate that intra-urban variability increases in the following order: regional sources (e.g., secondary SO_4^{2-} originating from EGUs) < area sources (e.g., on-road mobile sources) < point sources (e.g., metals from stacks of smelters).

Although limited information was available for $PM_{10-2.5}$, it does indicate a similar ordering, but without a regional component (resulting from the short lifetime of $PM_{10-2.5}$ compared to transport times on the regional scale). More discussion on source contributions to PM is available in Section 3.6.

2.1.7. Policy-Relevant Background

The background concentrations of PM that are useful for risk and policy assessments, which inform decisions about the NAAQS are referred to as policy-relevant background (PRB) concentrations. PRB concentrations have historically been defined by EPA as those concentrations that would occur in the U.S. in the absence of anthropogenic emissions in continental North America defined here as the U.S., Canada, and Mexico. For this document, PRB concentrations include contributions from natural sources everywhere in the world and from anthropogenic sources outside continental North America. Background concentrations so defined facilitated separation of pollution that can be controlled by U.S. regulations or through international agreements with neighboring countries from those that were judged to be generally uncontrollable by the U.S. Over time, consideration of potential broader ranging international agreements may lead to alternative determinations of which PM source contributions should be considered by EPA as part of PRB.

Contributions to PRB concentrations of PM include both primary and secondary natural and anthropogenic components. For this document, PRB concentrations of $PM_{2.5}$ for the continental U.S. were estimated using EPA's Community Multi-scale Air Quality (CMAQ) modeling system, a deterministic, chemical-transport model (CTM), using output from GEOS-Chem a global-scale model for CMAQ boundary conditions. PRB concentrations of $PM_{2.5}$ were estimated to be less than

$1 \mu\text{g}/\text{m}^3$ on an annual basis, with maximum daily average values in a range from 3.1 to $20 \mu\text{g}/\text{m}^3$ and having a peak of $63 \mu\text{g}/\text{m}^3$ at the nine national park sites across the U.S. used to evaluate model performance for this analysis. A description of the models and evaluation of their performance is given in Section 3.6 and further details about the calculations of PRB concentrations are given in Section 3.7.

2.2. Human Exposure

This section summarizes the findings from the recent exposure assessment literature. This summary is intended to support the interpretation of the findings from epidemiologic studies and reflects the material presented in Section 3.8. Attention is given to how concentration metrics can be used in exposure assessment and what errors and uncertainties are incurred for different approaches. Understanding of exposure errors is important because exposure error can potentially bias an estimate of a health effect or increase the size of confidence intervals around a health effect estimate.

2.2.1. Spatial Scales of PM Exposure Assessment

Assessing population-level exposure at the urban scale is particularly relevant for time-series epidemiologic studies, which provide information on the relationship between health effects and community-average exposure, rather than an individual's exposure. PM concentrations measured at a central-site ambient monitor are used as surrogates for personal PM exposure. However, the correlation between the PM concentration measured at central-site ambient monitor(s) and the unknown true community average concentration depends on the spatial distribution of PM, the location of the monitoring site(s) chosen to represent the community average, and division of the community by terrain features or local sources into several sub-communities that differ in the temporal pattern of pollution. Concentrations of SO_4^{2-} and some components of SOA measured at central-site monitors are expected to be uniform in urban areas because of the regional nature of their sources. However, this is not true for primary components like EC whose sources are strongly spatially variable in urban areas.

At micro-to-neighborhood scales, heterogeneity of sources and topography contribute to variability in exposure. This is particularly true for $PM_{10-2.5}$ and for UFPs, which have spatially

variable urban sources and loss processes (mainly gravitational settling for $PM_{10-2.5}$ and coagulation for UFPs) that also limit their transport from sources more readily than for $PM_{2.5}$. Personal activity patterns also vary across urban areas and across regions. Some studies, conducted mainly in Europe, have found personal $PM_{2.5}$ and PM_{10} exposures for pedestrians in street canyons to be higher than ambient concentrations measured by urban central site ambient monitors. Likewise, microenvironmental UFP concentrations were observed to be substantially higher in near-road environments, street canyons, and tunnels when compared with urban background concentrations.

In-vehicle UFP and $PM_{2.5}$ exposures can also be important. As a result, concentrations measured by ambient monitors likely do not reflect the contributions of UFP or $PM_{2.5}$ exposures to individuals while commuting. There is significant variability within and across regions of the country with respect to indoor exposures to ambient PM. Infiltrated ambient PM concentrations depend in part on the ventilation properties of the building or vehicle in which the person is exposed. PM infiltration factors depend on particle size, chemical composition, season, and region of the country. Infiltration can best be modeled dynamically rather than being represented by a single value. Season is important to PM infiltration because it affects the ventilation practices (e.g., open windows) used. In addition, ambient temperature and humidity conditions affect the transport, dispersion, and size distribution of PM. Residential air exchange rates have been observed to be higher in the summer for regions with low air conditioning usage. Regional differences in air exchange rates (Southwest < Southeast < Northeast < Northwest) also reflect ventilation practices. Differential infiltration occurs as a function of PM size and composition (the latter of which is described below). PM infiltration is larger for accumulation mode particles than for UFPs and $PM_{10-2.5}$. Differential infiltration by size fraction can affect exposure estimates if not accurately characterized.

2.2.2. Exposure to PM Components and Copollutants

Emission inventories and source apportionment studies suggest that sources of PM exposure vary by region. Comparison of studies performed in the eastern U.S. with studies performed in the western U.S. suggest that the contribution of SO_4^{2-} to exposure is higher for the East (16-46%) compared with the West (~4%) and that motor vehicle emissions and secondary NO_3^- are larger sources of exposure for the West (~9%) as compared with the East (~4%). Results of source apportionment studies of exposure to SO_4^{2-} indicate that SO_4^{2-} exposures are mainly attributable to ambient sources. Source apportionment for OC and EC is difficult because they originate from both indoor and outdoor sources. Exposure to OC of indoor and outdoor origin can be distinguished by the presence of aliphatic C-H groups generated indoors, since outdoor concentrations of aliphatic

C-H are low. Studies of personal exposure to ambient trace metal have shown significant variation among cities and over seasons. This is in response to geographic and seasonal variability in sources including incinerator operation, fossil fuel combustion, biomass combustion (wildfires), and the resuspension of crustal materials in the built environment. Differential infiltration is also affected by variations in particle composition and volatility. For example, EC infiltrates more readily than OC. This can lead to outdoor-indoor differentials in PM composition.

Some studies have explored the relationship between PM and copollutant gases and suggested that certain gases can serve as surrogates for describing exposure to other air pollutants. The findings indicate that ambient concentrations of gaseous copollutants can act as surrogates for personal exposure to ambient PM. Several studies have concluded that ambient concentrations of O_3 , NO_2 , and SO_2 are associated with the ambient component of personal exposure to total $PM_{2.5}$. If associations between ambient gases and personal exposure to $PM_{2.5}$ of ambient origin exist, such associations are complex and vary by season and location.

2.2.3. Implications for Epidemiologic Studies

In epidemiologic studies, exposure may be estimated using various approaches, most of which rely on measurements obtained using central site monitors. The magnitude and direction of the biases introduced through error in exposure measurement depend on the extent to which the error is associated with the measured PM concentration. In general, when exposure error is not strongly correlated with the measured PM concentration, bias is toward the null and effect estimates are

underestimated. Moreover, lack of information regarding exposure measurement error can also add uncertainty to the health effects estimate.

One important factor to be considered is the spatial variation in PM concentrations. The degree of urban-scale spatial variability in PM concentrations varies across the country and by size fraction. PM_{2.5} concentrations are relatively well-correlated across monitors in the urban areas examined for this assessment. The limited available evidence indicates that there is greater spatial variability in

PM_{10-2.5} concentrations than PM_{2.5} concentrations, resulting in increased exposure error for the larger size fraction. Likewise, studies have shown UFPs to be more spatially variable across urban areas compared to PM_{2.5}. Even if PM_{2.5}, PM_{10-2.5}, or UFP concentrations measured at sites within an urban area are generally highly correlated, significant spatial variation in their concentrations can occur on any given day. In addition, there can be differential exposure errors for PM components (e.g., SO₄²⁻, OC, EC). Current information suggests that UFPs, PM_{10-2.5}, and some PM components are more spatially variable than PM_{2.5}. Spatial variability of these PM indicators adds uncertainty to exposure estimates.

Overall, recent studies generally confirm and build upon the key conclusions of the 2004 PM AQCD: separation of total PM exposures into ambient and nonambient components reduces potential uncertainties in the analysis and interpretation of PM health effects data; and ambient PM concentration can be used as a surrogate for ambient PM exposure in community time-series epidemiologic studies because the change in ambient PM concentration should be reflected in the

change in the health risk coefficient. The use of the community average ambient PM_{2.5} concentration as a surrogate for the community average personal exposure to ambient PM_{2.5} is not expected to change the principal conclusions from time-series and most panel epidemiologic studies that use community average health and pollution data. Several recent studies support this by showing how the ambient component of personal exposure to PM_{2.5} could be estimated using various tracer and source apportionment techniques and by showing that the ambient component is highly correlated with ambient concentrations of PM_{2.5}. These studies show that the non-ambient component of personal exposure to PM_{2.5} is largely uncorrelated with ambient PM_{2.5} concentrations. A few panel epidemiologic studies have included personal as well as ambient monitoring data, and generally reported associations with all types of PM measurements. Epidemiologic studies of long-term exposure typically exploit the differences in PM concentration across space, as well as time, to estimate the effect of PM on the health outcome of interest. Long-term exposure estimates are most accurate for pollutants that do not vary substantially within the geographic area studied.

2.3. Health Effects

This section evaluates the evidence from toxicological, controlled human exposure, and epidemiologic studies that examined the health effects associated with short- and long-term exposure to PM (i.e., PM_{2.5}, PM_{10-2.5} and UFPs). The results from the health studies evaluated in combination with the evidence from atmospheric chemistry and exposure assessment studies contribute to the causal determinations made for the health outcomes discussed in this assessment (a description of the causal framework can be found in Section 1.5.4). In the following sections a discussion of the causal determinations will be presented by PM size fraction and exposure duration (i.e., short- or long-term exposure) for the health effects for which sufficient evidence was available to conclude a causal, likely to be causal or suggestive relationship. Although not presented in depth in this chapter, a detailed discussion of the underlying evidence used to formulate each causal determination can be found in Chapters 6 and 7.

2.3.1. Exposure to PM_{2.5}

2.3.1.1. Effects of Short-Term Exposure to PM_{2.5}

Table 2-1. Summary of causal determinations for short-term exposure to PM_{2.5}.

Size Fraction	Outcome	Causality Determination
PM _{2.5}	Cardiovascular Effects	Causal
	Respiratory Effects	Likely to be causal
Mortality		Causal

Cardiovascular Effects

Epidemiologic studies that examined the effect of PM_{2.5} on cardiovascular emergency department (ED) visits and hospital admissions reported consistent positive associations (predominantly for ischemic heart disease [IHD] and congestive heart failure [CHF]), with the majority of studies reporting increases ranging from 0.5 to 3.4% per 10 µg/m³ increase in PM_{2.5}. These effects were observed in study locations with mean¹ 24-h avg PM_{2.5} concentrations ranging from 7-18 µg/m³ (Section 6.2.10). The largest U.S.-based multicity study evaluated, Medicare Air Pollution Study (MCAPS), provided evidence of regional heterogeneity (e.g., the largest excess risks occurred in the Northeast [1.08%]) and seasonal variation (e.g., the largest excess risks occurred during the winter season [1.49%]) in PM_{2.5} cardiovascular disease (CVD) risk estimates, which is consistent with the null findings of several single-city studies conducted in the western U.S. These associations are supported by multicity epidemiologic studies that observed consistent positive associations between short-term exposure to PM_{2.5} and cardiovascular mortality and also reported regional and seasonal variability in risk estimates. The multicity studies evaluated reported consistent increases in cardiovascular mortality ranging from 0.47 to 0.85% in study locations with mean 24-h avg PM_{2.5} concentrations above 12.8 µg/m³ (Table 6-15).

Controlled human exposure studies have demonstrated PM_{2.5}-induced changes in various measures of cardiovascular function among healthy and health-compromised adults. The most consistent evidence is for altered vasomotor function following exposure to diesel exhaust (DE) or CAPs with O₃ (Section 6.2.4.2). Although these findings provide biological plausibility for the observations from epidemiologic studies, the fresh DE used in the controlled human exposure studies evaluated contains gaseous components (e.g., CO, NO_x), and therefore, the possibility that some of the changes in vasomotor function might be due to gaseous components cannot be ruled out. Furthermore, the prevalence of UFPs in fresh DE limits the ability to conclusively attribute the observed effects to either the UF fraction or PM_{2.5} as a whole. An evaluation of toxicological studies found evidence for altered vessel tone and microvascular reactivity, which provide coherence and biological plausibility for the vasomotor effects that have been observed in both the controlled human exposure and epidemiologic studies (Section 6.2.4.3). However, most of these toxicological studies exposed animals via intratracheal (IT) instillation or using relatively high inhalation concentrations. In addition to the effects observed on vasomotor function, myocardial ischemia has been observed across disciplines through PM_{2.5} effects on ST-segment depression, with toxicological studies providing biological plausibility by demonstrating reduced blood flow during ischemia (Section 6.2.3). There is also a growing body of evidence from controlled human exposure and toxicological studies demonstrating PM_{2.5}-induced changes on heart rate variability (HRV) and

¹ In this context mean represents the arithmetic mean of 24-h avg PM concentrations.

markers of systemic oxidative stress (Sections 6.2.1 and 6.2.9, respectively). Additional but inconsistent effects of PM_{2.5} on blood pressure (BP), blood coagulation markers, and markers of systemic inflammation have also been reported across disciplines. Toxicological studies have provided biologically plausible mechanisms (e.g., increased right ventricular pressure and diminished cardiac contractility) for the associations observed between PM_{2.5} and CHF in epidemiologic studies. Together, the collective evidence from epidemiologic, controlled human exposure, and toxicological studies is sufficient to conclude that **a causal relationship exists between short-term exposures to PM_{2.5} and cardiovascular effects.**

Respiratory Effects

The recent epidemiologic studies evaluated report consistent positive associations between short-term exposure to PM_{2.5} and respiratory ED visits and hospital admissions for chronic obstructive pulmonary disease (COPD) and respiratory infections (Section 6.3). Positive associations were also observed for asthma ED visits and hospital admissions for adults and children combined, but effect estimates are imprecise and not consistently positive for children alone. Most studies reported effects in the range of ~1% to 4% increase in respiratory hospital admissions and ED visits and were observed in study locations with mean 24-h avg PM_{2.5} concentrations ranging from 6.1-22 µg/m³. Additionally, multicity epidemiologic studies reported consistent positive associations between short-term exposure to PM_{2.5} and respiratory mortality as well as regional and seasonal variability in risk estimates. The multicity studies evaluated reported consistent, precise increases in respiratory mortality ranging from 1.67 to 2.20% in study locations with mean 24-h avg PM_{2.5} concentrations above 12.8 µg/m³ (Table 6-15). Evidence for PM_{2.5}-related respiratory effects was also observed in panel studies, which indicate associations with respiratory symptoms, pulmonary function, and pulmonary inflammation among asthmatic children. Although not consistently observed, some controlled human exposure studies have reported small decrements in various measures of pulmonary function following controlled exposures to PM_{2.5} (Section 6.3.2.2). Controlled human exposure studies using adult volunteers have demonstrated increased markers of pulmonary inflammation following exposure to a variety of different particle types; oxidative responses to DE and wood smoke; and exacerbations of allergic responses and allergic sensitization following exposure to DE particles (Section 6.3). Toxicological studies have provided additional support for PM_{2.5}-related respiratory effects through inhalation exposures of animals to CAPs, DE, other traffic-related PM and wood smoke. These studies reported an array of respiratory effects including altered pulmonary function, mild pulmonary inflammation and injury, oxidative responses, airway hyperresponsiveness (AHR) in allergic and non-allergic animals, exacerbations of allergic responses, and increased susceptibility to infections (Section 6.3). Overall, the evidence for an effect of PM_{2.5} on respiratory outcomes is somewhat restricted by limited coherence between some of the findings from epidemiologic and controlled human exposure studies for the specific health outcomes reported and the sub-populations in which those health outcomes occur. Epidemiologic studies have reported variable results among specific respiratory outcomes, specifically in asthmatics (e.g., increased respiratory symptoms in asthmatic children, but not increased asthma hospital admissions and ED visits) (Section 6.3.8). Additionally, respiratory effects have not been consistently demonstrated following controlled exposures to PM_{2.5} among asthmatics or individuals with COPD. Collectively, the epidemiologic, controlled human exposure, and toxicological studies evaluated demonstrate a wide range of respiratory responses, and although results are not fully consistent and coherent across studies the evidence is sufficient to conclude that **a causal relationship is likely to exist between short-term exposures to PM_{2.5} and respiratory effects.**

Mortality

An evaluation of the epidemiologic literature indicates consistent positive associations between short-term exposure to PM_{2.5} and all-cause, cardiovascular-, and respiratory-related mortality (Section 6.5.2.2.). The evaluation of multicity studies found that consistent and precise risk estimates for all-cause (nonaccidental) mortality that ranged from 0.29 to 1.21% per 10 µg/m³

increase in PM_{2.5} at lags of 1 and 0-1 days. In these study locations, mean 24-h avg PM_{2.5} concentrations were 12.8 µg/m³ and above (Table 6-15). Cardiovascular-related mortality risk estimates were found to be similar to those for all-cause mortality; whereas, the risk estimates for respiratory-related mortality were consistently larger (i.e., 1.01-2.2%) using the same lag periods and averaging indices. The studies evaluated that examined the relationship between short-term exposure to PM_{2.5} and cardiovascular effects (Section 6.2) provide coherence and biological plausibility for PM_{2.5}-induced cardiovascular mortality, which represents the largest component of total (nonaccidental) mortality (~ 35%) (American Heart Association, 2009, 198920). However, as noted in Section 6.3, there is limited coherence between some of the respiratory morbidity findings from epidemiologic and controlled human exposure studies for the specific health outcomes reported and the subpopulations in which those health outcomes occur, complicating the interpretation of the PM_{2.5} respiratory mortality effects observed. Regional and seasonal patterns in PM_{2.5} risk estimates were observed with the greatest effect estimates occurring in the eastern U.S. and during the spring. Of the studies evaluated only Burnett et al. (2004, 086247), a Canadian multicity study, analyzed gaseous pollutants and found mixed results, with possible confounding of PM_{2.5} risk estimates by NO₂. Although the recently evaluated U.S.-based multicity studies did not analyze potential confounding of PM_{2.5} risk estimates by gaseous pollutants, evidence from the limited number of single-city studies evaluated in the 2004 PM AQCD (U.S. EPA, 2004, 056905) suggest that gaseous copollutants do not confound the PM_{2.5}-mortality association. This is further supported by studies that examined the PM₁₀-mortality relationship. An examination of effect modifiers (e.g., demographic and socioeconomic factors), specifically air conditioning use as an indicator for decreased pollutant penetration indoors, has suggested that PM_{2.5} risk estimates increase as the percent of the population with access to air conditioning decreases. Collectively, the epidemiologic literature provides evidence that **a causal relationship exists between short-term exposures to PM_{2.5} and mortality.**

2.3.1.2. Effects of Long-Term Exposure to PM_{2.5}

Table 2-2. Summary of causal determinations for long-term exposure to PM_{2.5}.

Size Fraction	Outcome	Causality Determination
<u>Cardiovascular Effects</u>		<u>Causal</u>
	<u>Respiratory Effects</u>	<u>Likely to be causal</u>
	<u>PM_{2.5} Mortality</u>	<u>Causal</u>
	<u>Reproductive and Developmental</u>	<u>Suggestive</u>
<u>Cancer, Mutagenicity, and Genotoxicity</u>		<u>Suggestive</u>

Cardiovascular Effects

The strongest evidence for cardiovascular health effects related to long-term exposure to PM_{2.5} comes from large, multicity U.S.-based studies, which provide consistent evidence of an association between long-term exposure to PM_{2.5} and cardiovascular mortality (Section 7.2.10). These associations are supported by a large U.S.-based epidemiologic study (i.e., Women’s Health Initiative [WHI] study) that reports associations between PM_{2.5} and CVDs among post-menopausal women using a 1-yr avg PM_{2.5} concentration (mean = 13.5 µg/m³) (Section 7.2). However, epidemiologic studies that examined subclinical markers of CVD report inconsistent findings. Epidemiologic studies have also provided some evidence for potential modification of the PM_{2.5}-CVD association when examining individual-level data, specifically smoking status and the use of anti-

hyperlipidemics. Although epidemiologic studies have not consistently detected effects on markers of atherosclerosis due to long-term exposure to $PM_{2.5}$, toxicological studies have provided strong evidence for accelerated development of atherosclerosis in ApoE^{-/-} mice exposed to CAPs and have shown effects on coagulation, experimentally-induced hypertension, and vascular reactivity (Section 7.2.1.2). Evidence from toxicological studies provides biological plausibility and coherence with studies of short-term exposure and cardiovascular morbidity and mortality, as well as with studies that examined long-term exposure to $PM_{2.5}$ and cardiovascular mortality. Taken together, the evidence from epidemiologic and toxicological studies is sufficient to conclude that **a causal relationship exists between long-term exposures to $PM_{2.5}$ and cardiovascular effects.**

Respiratory Effects

Recent epidemiologic studies conducted in the U.S. and abroad provide evidence of associations between long-term exposure to $PM_{2.5}$ and decrements in lung function growth, increased respiratory symptoms, and asthma development in study locations with mean $PM_{2.5}$ concentrations ranging from 13.8 to 30 $\mu\text{g}/\text{m}^3$ during the study periods (Section 7.3.1.1 and Section 7.3.2.1). These results are supported by studies that observed associations between long-term exposure to PM_{10} and an increase in respiratory symptoms and reductions in lung function growth in areas where PM_{10} is dominated by $PM_{2.5}$. However, the evidence to support an association with long-term exposure to $PM_{2.5}$ and respiratory mortality is limited (Figure 7-7). Subchronic and chronic toxicological studies of CAPs, DE, roadway air and woodsmoke provide coherence and biological plausibility for the effects observed in the epidemiologic studies. These toxicological studies have presented some evidence for altered pulmonary function, mild inflammation, oxidative responses, immune suppression, and histopathological changes including mucus cell hyperplasia (Section 7.3). Exacerbated allergic responses have been demonstrated in animals exposed to DE and wood smoke. In addition, pre- and postnatal exposure to ambient levels of urban particles was found to affect lung development in an animal model. This finding is important because impaired lung development is one mechanism by which PM exposure may decrease lung function growth in children. Collectively, the evidence from epidemiologic and toxicological studies is sufficient to conclude that **a causal relationship is likely to exist between long-term exposures to $PM_{2.5}$ and respiratory effects.**

Mortality

The recent epidemiologic literature reports associations between long-term $PM_{2.5}$ exposure and increased risk of mortality. Mean $PM_{2.5}$ concentrations ranged from 13.2 to 29 $\mu\text{g}/\text{m}^3$ during the study period in these areas (Section 7.6). When evaluating cause-specific mortality, the strongest evidence can be found when examining associations between $PM_{2.5}$ and cardiovascular mortality, and positive associations were also reported between $PM_{2.5}$ and lung cancer mortality (Figure 7-7). The cardiovascular mortality association has been confirmed further by the extended Harvard Six Cities and American Cancer Society studies, which both report strong associations between long-term exposure to $PM_{2.5}$ and cardiopulmonary and IHD mortality (Figure 7-7). Additional new evidence from a study that used the WHI cohort found a particularly strong association between long-term exposure to $PM_{2.5}$ and CVD mortality in post-menopausal women. Fewer studies have evaluated the respiratory component of cardiopulmonary mortality, and, as a result, the evidence to support an association with long-term exposure to $PM_{2.5}$ and respiratory mortality is limited (Figure 7-7). The evidence for cardiovascular and respiratory morbidity due to short- and long-term exposure to $PM_{2.5}$ provides biological plausibility for cardiovascular- and respiratory-related mortality. Collectively, the evidence is sufficient to conclude that **a causal relationship exists between long-term exposures to $PM_{2.5}$ and mortality.**

Reproductive and Developmental Effects

Evidence is accumulating for PM_{2.5} effects on low birth weight and infant mortality, especially due to respiratory causes during the post-neonatal period. The mean PM_{2.5} concentrations during the study periods ranged from 5.3-27.4 µg/m³ (Section 7.4), with effects becoming more precise and consistently positive in locations with mean PM_{2.5} concentrations of 15 µg/m³ and above (Section 7.4). Exposure to PM_{2.5} was usually associated with greater reductions in birth weight than exposure to PM₁₀. The evidence from a few U.S. studies that investigated PM₁₀ effects on fetal growth, which reported similar decrements in birth weight, provide consistency for the PM_{2.5} associations observed and strengthen the interpretation that particle exposure may be causally related to reductions in birth weight. The epidemiologic literature does not consistently report associations between long-term exposure to PM and preterm birth, growth restriction, birth defects or decreased sperm quality. Toxicological evidence supports an association between PM_{2.5} and PM₁₀ exposure and adverse reproductive and developmental outcomes, but provide little mechanistic information or biological plausibility for an association between long-term PM exposure and adverse birth outcomes (e.g., low birth weight or infant mortality). New evidence from animal toxicological studies on heritable mutations is of great interest, and warrants further investigation. Overall, the epidemiologic and toxicological evidence is **suggestive of a causal relationship between long-term exposures to PM_{2.5} and reproductive and developmental outcomes.**

Cancer, Mutagenicity, and Genotoxicity

Multiple epidemiologic studies have shown a consistent positive association between PM_{2.5} and lung cancer mortality, but studies have generally not reported associations between PM_{2.5} and lung cancer incidence (Section 7.5). Animal toxicological studies have examined the potential relationship between PM and cancer, but have not focused on specific size fractions of PM. Instead they have examined ambient PM, wood smoke, and DEP. A number of studies indicate that ambient urban PM, emissions from wood/biomass burning, emissions from coal combustion, and gasoline and DE are mutagenic, and that PAHs are genotoxic. These findings are consistent with earlier studies that concluded that ambient PM and PM from specific combustion sources are mutagenic and genotoxic and provide biological plausibility for the results observed in the epidemiologic studies. A limited number of epidemiologic and toxicological studies examined epigenetic effects, and demonstrate that PM induces some changes in methylation. However, it has yet to be determined **how these alterations in the genome could influence the initiation and promotion of cancer. Additionally, inflammation and immune suppression induced by exposure to PM may confer susceptibility to cancer. Collectively, the evidence from epidemiologic studies, primarily those of lung cancer mortality, along with the toxicological studies that show some evidence of the mutagenic and genotoxic effects of PM is suggestive of a causal relationship between long-term exposures to PM_{2.5} and cancer.**

2.3.2. Integration of PM_{2.5} Health Effects

In epidemiologic studies, short-term exposure to PM_{2.5} is associated with a broad range of respiratory and cardiovascular effects, as well as mortality. For cardiovascular effects and mortality, the evidence supports the existence of a causal relationship with short-term PM_{2.5} exposure; while the evidence indicates that a causal relationship is likely to exist between short-term PM_{2.5} exposure and respiratory effects. The effect estimates from recent and older U.S. and Canadian-based epidemiologic studies that examined the relationship between short-term exposure to PM_{2.5} and health outcomes with mean 24-h avg PM_{2.5} concentrations <17 µg/m³ are shown in Figure 2-1. **A number of different health effects are included in Figure 2-1 to provide an integration of the range of effects by mean concentration, with a focus on cardiovascular and respiratory effects and all-cause (nonaccidental) mortality (i.e., health effects categories with at least a suggestive causal determination). A pattern of consistent positive associations with mortality and morbidity effects can be seen in this figure. Mean PM_{2.5} concentrations ranged from 6.1 to 16.8 µg/m³ in these study locations.**

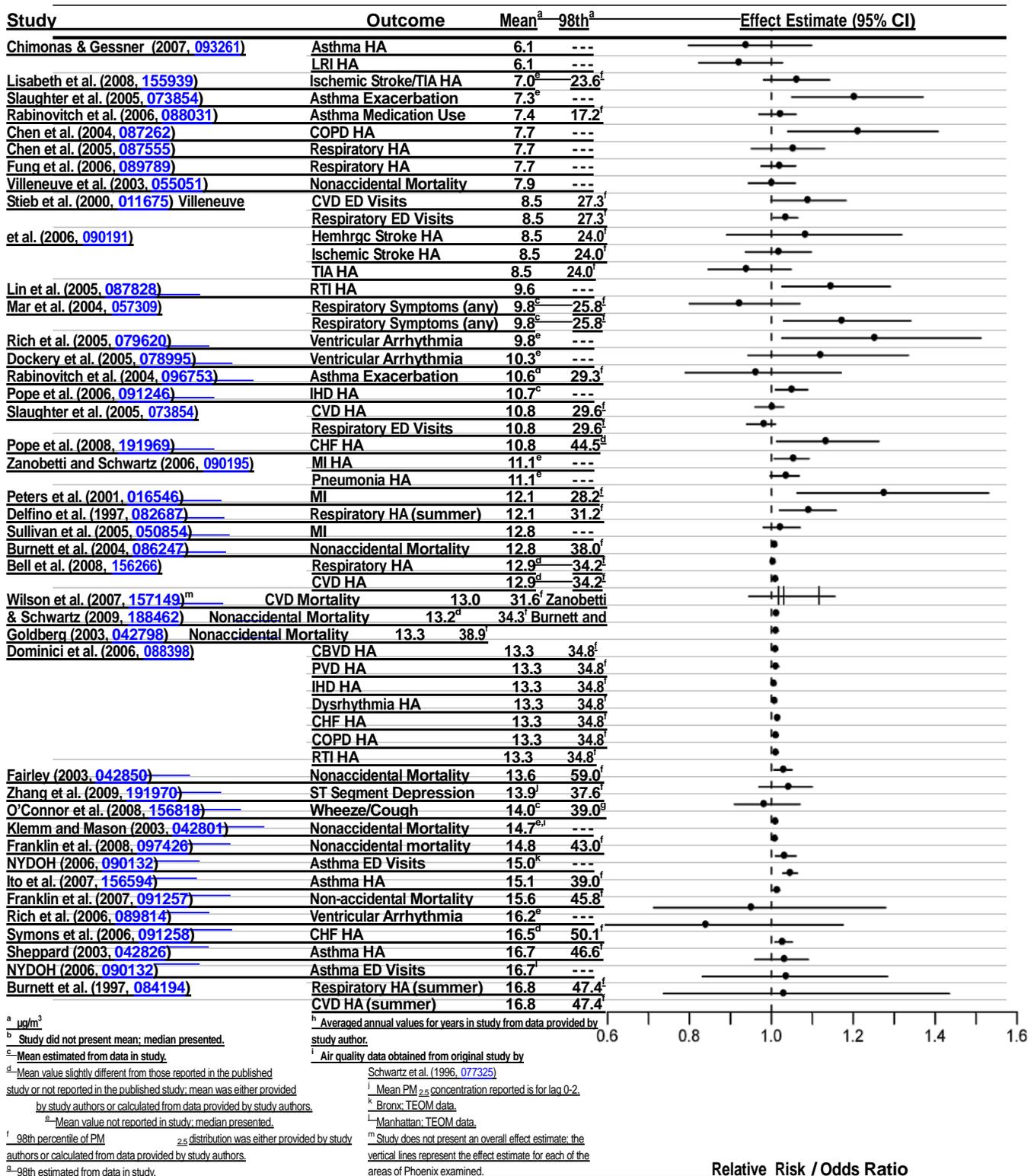


Figure 2-1. Summary of effect estimates (per 10 $\mu\text{g}/\text{m}^3$) by increasing concentration from U.S. studies examining the association between short-term exposure to PM_{2.5} and cardiovascular and respiratory effects, and mortality, conducted in locations where the reported mean 24-h avg PM_{2.5} concentrations were <17 $\mu\text{g}/\text{m}^3$.

Long-term exposure to PM_{2.5} has been associated with health outcomes similar to those found in the short-term exposure studies, specifically for respiratory and cardiovascular effects and mortality. As found for short-term PM_{2.5} exposure, the evidence indicates that a causal relationship exists between long-term PM_{2.5} exposure and cardiovascular effects and mortality, and that a causal relationship is likely to exist between long-term PM_{2.5} exposure and effects on the respiratory system.

Figure 2-2 highlights the findings of epidemiologic studies where the long-term mean PM_{2.5} concentrations were < 29 µg/m³. A range of health outcomes are displayed (including cardiovascular mortality, all-cause mortality, infant mortality, and bronchitis) ordered by mean concentration. The range of mean PM_{2.5} concentrations in these studies was 10.7-29 µg/m³ during the study periods. Additional studies not included in this figure that focus on subclinical outcomes, such as changes in lung function or atherosclerotic markers also report effects in areas with similar concentrations (Sections 7.2 and 7.3). Although not highlighted in the summary figure, long-term PM_{2.5} exposure studies also provide evidence for reproductive and developmental effects (i.e., low birth weight) and cancer (i.e., lung cancer mortality) in response to exposure to PM_{2.5}.

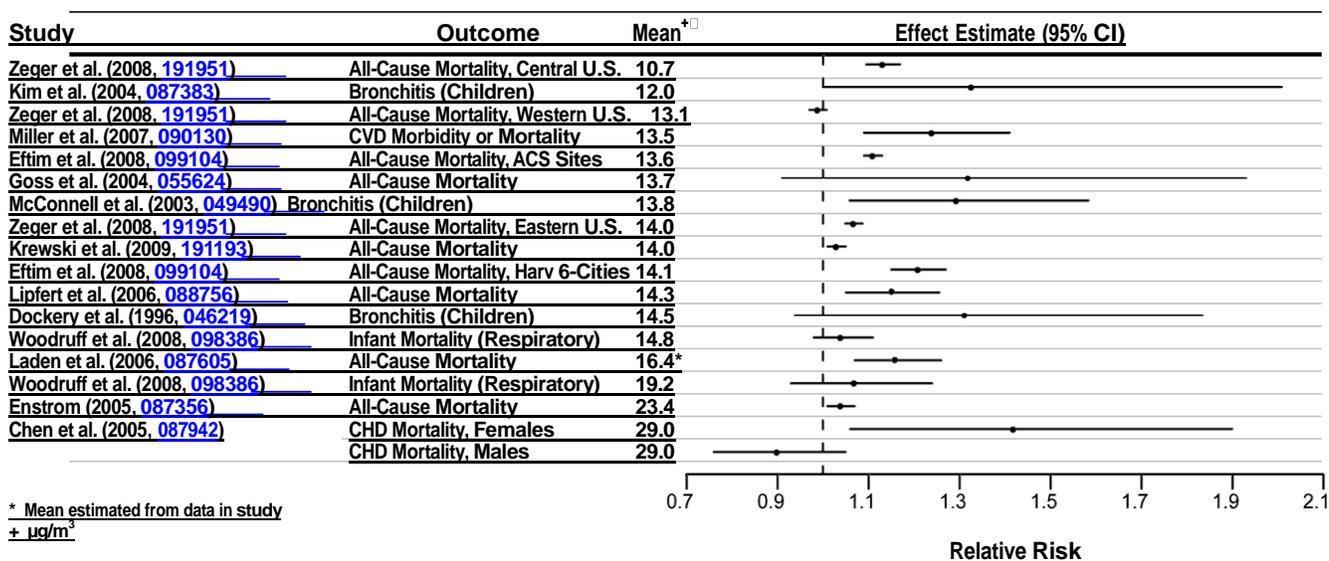


Figure 2-2. Summary of effect estimates (per 10 µg/m³) by increasing concentration from U.S. studies examining the association between long-term exposure to PM_{2.5} and cardiovascular and respiratory effects, and mortality.

The observations from both the short- and long-term exposure studies are supported by experimental findings of PM_{2.5}-induced subclinical and clinical cardiovascular effects. Epidemiologic studies have shown an increase in ED visits and hospital admissions for IHD upon exposure to PM_{2.5}. These effects are coherent with the changes in vasomotor function and ST-segment depression observed in both toxicological and controlled human exposure studies. It has been postulated that exposure to PM_{2.5} can lead to myocardial ischemia through an effect on the autonomic nervous system or by altering vasomotor function. PM-induced systemic inflammation, oxidative stress and/or endothelial dysfunction may contribute to altered vasomotor function. These effects have been demonstrated in recent animal toxicological studies, along with altered microvascular reactivity, altered vessel tone, and reduced blood flow during ischemia. Toxicological studies demonstrating increased right ventricular pressure and diminished cardiac contractility also provide biological plausibility for the associations observed between PM_{2.5} and CHF in epidemiologic studies. Thus, the overall evidence from the short-term epidemiologic, controlled human exposure, and toxicological studies evaluated provide coherence and biological plausibility for cardiovascular effects related to myocardial ischemia and CHF. Coherence in the cardiovascular effects observed

can be found in long-term exposure studies, especially for CVDs among post-menopausal women. Additional studies provide limited evidence for subclinical measures of atherosclerosis in epidemiologic studies with stronger evidence from toxicological studies that have demonstrated accelerated development of atherosclerosis in ApoE^{-/-} mice exposed to PM_{2.5} CAPs along with effects on coagulation, experimentally-induced hypertension, and vascular reactivity. Repeated acute responses to PM may lead to cumulative effects that manifest as chronic disease, such as atherosclerosis. Contributing factors to atherosclerosis development include systemic inflammation, endothelial dysfunction, and oxidative stress all of which are associated with PM_{2.5} exposure. However, it has not yet been determined whether PM initiates or promotes atherosclerosis. The evidence from both short- and long-term exposure studies on cardiovascular morbidity provide coherence and biological plausibility for the cardiovascular mortality effects observed when examining both exposure durations. In addition, cardiovascular hospital admission and mortality studies that examined the PM₁₀ concentration-response relationship found evidence of a log-linear no-threshold relationship between PM exposure and cardiovascular-related morbidity (Section 6.2) and mortality (Section 6.5). Epidemiologic studies have also reported respiratory effects related to short-term exposure to PM_{2.5}, which include increased ED visits and hospital admissions, as well as alterations in lung function and respiratory symptoms in asthmatic children. These respiratory effects were found to be generally robust to the inclusion of gaseous pollutants in copollutant models with the strongest evidence from the higher powered studies (Figure 6-9 and Figure 6-15). Consistent positive associations were also reported between short-term exposure to PM_{2.5} and respiratory mortality in epidemiologic studies. However, uncertainties exist in the PM_{2.5}-respiratory mortality associations reported due to the limited number of studies that examined potential confounders of the PM_{2.5}-respiratory mortality relationship, and the limited information regarding the biological plausibility of the clinical and subclinical respiratory outcomes observed in the epidemiologic and controlled human exposure studies (Section 6.3) resulting in the progression to PM_{2.5}-induced respiratory mortality. Important new findings, which support the PM_{2.5}-induced respiratory effects mentioned above, include associations with post-neonatal (between 1 mo and 1 yr of age) respiratory mortality. Controlled human exposure studies provide some support for the respiratory findings from epidemiologic studies, with demonstrated increases in pulmonary inflammation following short-term exposure. However, there is limited and inconsistent evidence of effects in response to controlled exposures to PM_{2.5} on respiratory symptoms or pulmonary function among healthy adults or adults with respiratory disease. Long-term exposure epidemiologic studies provide additional evidence for PM_{2.5}-induced respiratory morbidity, but little evidence for an association with respiratory mortality. These epidemiologic morbidity studies have found decrements in lung function growth, as well as increased respiratory symptoms, and asthma. Toxicological studies provide coherence and biological plausibility for the respiratory effects observed in response to short and long-term exposures to PM by demonstrating a wide array of biological responses including: altered pulmonary function, mild pulmonary inflammation and injury, oxidative responses, and histopathological changes in animals exposed by inhalation to PM_{2.5} derived from a wide variety of sources. In some cases, prolonged exposures led to adaptive responses. Important evidence was also found in an animal model for altered lung development following pre- and post-natal exposure to urban air, which may provide a mechanism to explain the reduction in lung function growth observed in children in response to long-term exposure to PM.

Additional respiratory-related effects have been tied to allergic responses. Epidemiologic studies have provided evidence for increased hospital admissions for allergic symptoms (e.g., allergic rhinitis) in response to short- and long-term exposure to PM_{2.5}. Panel studies also positively associate long-term exposure to PM_{2.5} and PM₁₀ with indicators of allergic sensitization. Controlled human exposure and toxicological studies provide coherence for the exacerbation of allergic symptoms, by showing that PM_{2.5} can promote allergic responses and intensify existing allergies. Allergic responses require repeated exposures to antigen over time and co-exposure to an adjuvant (possibly DE particles or UF CAPs) can enhance this response. Allergic sensitization often underlies allergic asthma, characterized by inflammation and AHR. In this way, repeated or chronic exposures involving multifactorial responses (immune system activation, oxidative stress, inflammation) can lead to irreversible outcomes. Epidemiologic studies have also reported evidence for increased hospital admissions for respiratory infections in response to both short- and long-term exposures to PM_{2.5}. Toxicological studies suggest that PM impairs innate immunity, which is the first line of

defense against infection, providing coherence for the respiratory infection effects observed in epidemiologic studies.

The difference in effects observed across studies and between cities may be attributed, at least in part, to the differences in PM composition across the U.S. Differences in PM toxicity may result from regionally varying PM composition and size distribution, which in turn reflects differences in sources and PM volatility. A person's exposure to ambient PM will also vary due to regional differences in personal activity patterns, microenvironmental characteristics and the spatial variability of PM concentrations in urban areas. Regional differences in PM_{2.5} composition are outlined briefly in Section 2.1 above and in more detail in Section 3.5. An examination of data from the CSN indicates that East-West gradients exist for a number of PM components. Specifically, SO₄²⁻ concentrations are higher in the East, OC constitutes a larger fraction of PM in the West, and NO₃⁻ concentrations are highest in the valleys of central California and during the winter in the Midwest. However, the available evidence and the limited amount of city-specific speciated PM_{2.5} data does not allow conclusions to be drawn that specifically differentiate effects of PM in different locations.

It remains a challenge to determine relationships between specific constituents, combinations of constituents, or sources of PM_{2.5} and the various health effects observed. Source apportionment studies of PM_{2.5} have attempted to decipher some of these relationships and in the process have identified associations between multiple sources and various respiratory and cardiovascular health effects, as well as mortality. Although different source apportionment methods have been used across these studies, the methods used have been evaluated and found generally to identify the same sources and associations between sources and health effects (Section 6.6). While uncertainty remains, it has been recognized that many sources and components of PM_{2.5} contribute to health effects. Overall, the results displayed in Table 6-18 indicate that many constituents of PM_{2.5} can be linked with multiple health effects, and the evidence is not yet sufficient to allow differentiation of those constituents or sources that are more closely related to specific health outcomes.

Variability in the associations observed across PM_{2.5} epidemiologic studies may be due in part to exposure error related to the use of county-level air quality data. Because western U.S. counties tend to be much larger and more topographically diverse than eastern U.S. counties, the day-to-day variations in concentration at one site, or even for the average of several sites, may not correlate well with the day-to-day variations in all parts of the county. For example, site-to-site correlations as a function of distance between sites (Section 3.5.1.2) fall off rapidly with distance in Los Angeles, but high correlations extend to larger distances in eastern cities such as Boston and Pittsburgh. These differences may be attributed to a number of factors including topography, the built environment, climate, source characteristics, ventilation usage, and personal activity patterns. For instance, regional differences in climate and infrastructure can affect time spent outdoors or indoors, air conditioning usage, and personal activity patterns. Characteristics of housing stock may also cause regional differences in effect estimates because new homes tend to have lower infiltration factors than older homes. Biases and uncertainties in exposure estimates resulting from these aspects can, in turn, cause bias and uncertainty in associated health effects estimates.

The new evidence reviewed in this ISA greatly expands upon the evidence available in the 2004 PM AQCD particularly in providing greater understanding of the underlying mechanisms for PM_{2.5} induced cardiovascular and respiratory effects for both short- and long-term exposures. Recent studies have provided new evidence linking long-term exposure to PM_{2.5} with cardiovascular outcomes that has expanded upon the continuum of effects ranging from the more subtle subclinical measures to cardiopulmonary mortality.

2.3.3. Exposure to PM_{10-2.5}

2.3.3.1. Effects of Short-Term Exposure to PM_{10-2.5}

Table 2-3. Summary of causal determinations for short-term exposure to PM_{10-2.5}.

Size Fraction	Outcome	Causality Determination
PM _{10-2.5}	Cardiovascular Effects	Suggestive
	Respiratory Effects	Suggestive
Mortality		Suggestive

Cardiovascular Effects

Generally positive associations were reported between short-term exposure to PM_{10-2.5} and hospital admissions or ED visits for cardiovascular causes. These results are supported by a large U.S. multicity study of older adults that reported PM_{10-2.5} associations with CVD hospital admissions, and only a slight reduction in the PM_{10-2.5} risk estimate when included in a copollutant model with PM_{2.5} (Section 6.2.10). The PM_{10-2.5} associations with cardiovascular hospital admissions and ED visits were observed in study locations with mean 24-h avg PM_{10-2.5} concentrations ranging from 7.4 to 13 µg/m³. These results are supported by the associations observed between PM_{10-2.5} and cardiovascular mortality in areas with 24-h avg PM_{10-2.5} concentrations ranging from 6.1-16.4 µg/m³ (Section 6.2.11). The results of the epidemiologic studies were further confirmed by studies that examined dust storm events, which contain high concentrations of crustal material, and found an increase in cardiovascular-related ED visits and hospital admissions. Additional epidemiologic studies have reported PM_{10-2.5} associations with other cardiovascular health effects including supraventricular ectopy and changes in HRV (Section 6.2.1.1). Although limited in number, studies of controlled human exposures provide some evidence to support the alterations in HRV observed in the epidemiologic studies (Section 6.2.1.2). The few toxicological studies that examined the effect of PM_{10-2.5} on cardiovascular health effects used IT instillation due to the technical challenges in exposing rodents via inhalation to PM_{10-2.5}, and, as a result, provide only limited evidence on the biological plausibility of PM_{10-2.5} induced cardiovascular effects. The potential for PM_{10-2.5} to elicit an effect is supported by dosimetry studies, which show that a large proportion of inhaled particles in the 3-6 micron (d_{ac}) range can reach and deposit in the lower respiratory tract, particularly the tracheobronchial (TB) airways (Figures 4-3 and 4-4). Collectively, the evidence from epidemiologic studies, along with the more limited evidence from controlled human exposure and toxicological studies **is suggestive of a causal relationship between short-term exposures to PM_{10-2.5} and cardiovascular effects.**

Respiratory Effects

A number of recent epidemiologic studies conducted in Canada and France found consistent, positive associations between respiratory ED visits and hospital admissions and short-term exposure to PM_{10-2.5} in studies with mean 24-h avg concentrations ranging from 5.6-16.2 µg/m³ (Section 6.3.8). In these studies, the strongest relationships were observed among children, with less consistent evidence for adults and older adults (i.e., ≥ 65). In a large multicity study of older adults, PM_{10-2.5} was positively associated with respiratory hospital admissions in both single and copollutant models with PM_{2.5}. In addition, a U.S.-based multicity study found evidence for an increase in respiratory mortality upon short-term exposure to PM_{10-2.5}, but these associations have not been consistently

observed in single-city studies (Section 6.3.9). A limited number of epidemiologic studies have focused on specific respiratory morbidity outcomes, and found no evidence of an association with lower respiratory symptoms, wheeze, and medication use (Section 6.3.1.1). While controlled human exposure studies have not observed an effect on lung function or respiratory symptoms in healthy or asthmatic adults in response to short-term exposure to PM_{10-2.5}, healthy volunteers have exhibited an increase in markers of pulmonary inflammation. Toxicological studies using inhalation exposures are still lacking, but pulmonary injury has been observed in animals after IT instillation exposure (Section 6.3.5.3). In some cases, PM_{10-2.5} was found to be more potent than PM_{2.5} and effects were not attributable to endotoxin. Both rural and urban PM_{10-2.5} have induced inflammation and injury responses in rats or mice exposed via IT instillation, making it difficult to distinguish the health effects of PM_{10-2.5} from different environments. Overall, epidemiologic studies, along with the limited number of controlled human exposure and toxicological studies that examined PM_{10-2.5} respiratory effects provide evidence that **is suggestive of a causal relationship between short-term exposures to PM_{10-2.5} and respiratory effects.**

Mortality

The majority of studies evaluated in this review provide some evidence for mortality associations with PM_{10-2.5} in areas with mean 24-h avg concentrations ranging from 6.1-16.4 µg/m³. However, uncertainty surrounds the PM_{10-2.5} associations reported in the studies evaluated due to the different methods used to estimate PM_{10-2.5} concentrations across studies (e.g., direct measurement of PM_{10-2.5} using dichotomous samplers, calculating the difference between PM₁₀ and PM_{2.5} concentrations). In addition, only a limited number of PM_{10-2.5} studies have investigated potential confounding by gaseous copollutants or the influence of model specification on PM_{10-2.5} risk estimates.

A new U.S.-based multicity study, which estimated PM_{10-2.5} concentrations by calculating the difference between the county-average PM₁₀ and PM_{2.5}, found associations between PM_{10-2.5} and mortality across the U.S., including evidence for regional variability in PM_{10-2.5} risk estimates (Section 6.5.2.3). Additionally, the U.S.-based multicity study provides preliminary evidence for greater effects occurring during the warmer months (i.e., spring and summer). A multicity Canadian study provides additional evidence for an association between short-term exposure to PM_{10-2.5} and mortality (Section 6.5.2.3). Although consistent positive associations have been observed across both multi- and single-city studies, more data are needed to adequately characterize the chemical and biological components that may modify the potential toxicity of PM_{10-2.5} and compare the different methods used to estimate exposure. Overall, the evidence evaluated **is suggestive of a causal relationship between short-term exposures to PM_{10-2.5} and mortality.**

2.3.4. Integration of PM_{10-2.5} Effects

Epidemiologic, controlled human exposure, and toxicological studies have provided evidence that is suggestive for relationships between short-term exposure to PM_{10-2.5} and cardiovascular effects, respiratory effects, and mortality. Conclusions regarding causation for the various health effects and outcomes were made for PM_{10-2.5} as a whole regardless of origin, since PM_{10-2.5}-related effects have been demonstrated for a number of different environments (e.g., cities reflecting a wide range of environmental conditions). Associations between short-term exposure to PM_{10-2.5} and cardiovascular and respiratory effects, and mortality have been observed in locations with mean PM_{10-2.5} concentrations ranging from 5.6 to 33.2 µg/m³, and maximum PM_{10-2.5} concentrations ranging from 24.6 to 418.0 µg/m³ (Figure 2-3). A number of different health effects are included in Figure 2-3 to provide an integration of the range of effects by mean concentration, with a focus on cardiovascular and respiratory effects, and mortality (i.e., health effects categories with at least a suggestive causal determination). To date, a sufficient amount of evidence does not exist in order to draw conclusions regarding the health effects and outcomes associated with long-term exposure to PM_{10-2.5}.

In epidemiologic studies, associations between short-term exposure to PM_{10-2.5} and cardiovascular outcomes (i.e., IHD hospital admissions, supraventricular ectopy, and changes in HRV) have been found that are similar in magnitude to those observed in PM_{2.5} studies. Controlled human exposure studies have also observed alterations in HRV, providing consistency and coherence

for the effects observed in the epidemiologic studies. To date, only a limited number of toxicological studies have been conducted to examine the effects of PM_{10-2.5} on cardiovascular effects. All of these studies involved IT instillation due to the technical challenges of using PM_{10-2.5} for rodent inhalation studies. As a result, the toxicological studies evaluated provide limited biological plausibility for the PM_{10-2.5} effects observed in the epidemiologic and controlled human exposure studies.

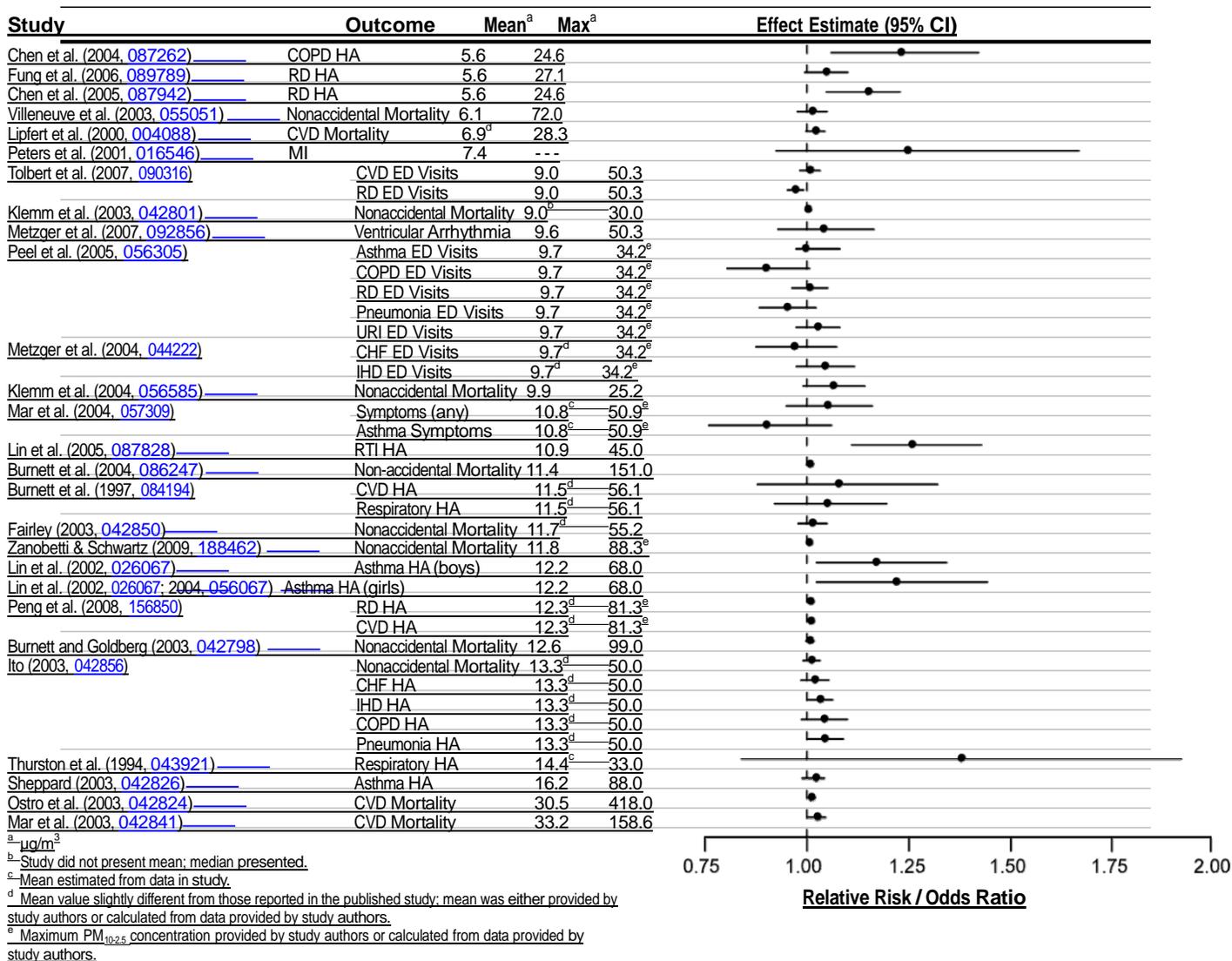


Figure 2-3. Summary of U.S. studies examining the association between short-term exposure to PM_{10-2.5} and cardiovascular morbidity/mortality and respiratory morbidity/mortality. All effect estimates have been standardized to reflect a 10 $\mu\text{g}/\text{m}^3$ increase in mean 24-h avg PM_{10-2.5} concentration and ordered by increasing concentration.

Limited evidence is available from epidemiologic studies for respiratory health effects and outcomes in response to short-term exposure to PM_{10-2.5}. An increase in respiratory hospital admissions and ED visits has been observed, but primarily in studies conducted in Canada and Europe. In addition, associations are not reported for lower respiratory symptoms, wheeze, or medication use. Controlled human exposure studies have not observed an effect on lung function or respiratory symptoms in healthy or asthmatic adults, but healthy volunteers have exhibited pulmonary inflammation. The toxicological studies (all IT instillation) provide evidence of

pulmonary injury and inflammation. In some cases, PM_{10-2.5} was found to be more potent than PM_{2.5} and effects were not solely attributable to endotoxin.

Currently, a national network is not in place to monitor PM_{10-2.5} concentrations. As a result, uncertainties surround the concentration at which the observed associations occur. Ambient concentrations of PM_{10-2.5} are generally determined by the subtraction of PM₁₀ and PM_{2.5} measurements, using various methods. For example, some epidemiologic studies estimate PM_{10-2.5} by taking the difference between collocated PM₁₀ and PM_{2.5} monitors while other studies have taken the difference between county average PM₁₀ and PM_{2.5} concentrations. Moreover, there are potential differences among operational flow rates and temperatures for PM₁₀ and PM_{2.5} monitors used to calculate PM_{10-2.5}. Therefore, there is greater error in ambient exposure to PM_{10-2.5} compared to PM_{2.5}. This would tend to increase uncertainty and make it more difficult to detect effects of PM_{10-2.5} in epidemiologic studies. In addition, the various differences between eastern and western U.S. counties can lead to exposure misclassification, and the potential underestimation of effects in western counties (as discussed for PM_{2.5} in Section 2.3.2).

It is also important to note that the chemical composition of PM_{10-2.5} can vary considerably by location, but city-specific speciated PM_{10-2.5} data are limited. PM_{10-2.5} may contain Fe, Si, Al, and base cations from soil, plant and insect fragments, pollen, fungal spores, bacteria, and viruses, as well as fly ash, brake lining particles, debris, and automobile tire fragments.

The 2004 PM AQCD presented the limited amount of evidence available that examined the potential association between exposure to PM_{10-2.5} and health effects and outcomes. The current evidence, primarily from epidemiologic studies, builds upon the results from the 2004 PM AQCD and indicates that short-term exposure to PM_{10-2.5} is associated with effects on both the cardiovascular and respiratory systems. However, variability in the chemical and biological composition of PM_{10-2.5}, limited evidence regarding effects of the various components of PM_{10-2.5}, and lack of clearly defined biological mechanisms for PM_{10-2.5}-related effects are important sources of uncertainty.

2.3.5. Exposure to UFPs

2.3.5.1. Effects of Short-Term Exposure to UFPs

Table 2-4. Summary of causal determinations for short-term exposure to UFPs.

<u>Size Fraction</u>	<u>Outcome</u>	<u>Causality Determination</u>
<u>UFPs</u>	<u>Cardiovascular Effects</u>	<u>Suggestive</u>
	<u>Respiratory Effects</u>	<u>Suggestive</u>

Cardiovascular Effects

Controlled human exposure studies provide the majority of the evidence for cardiovascular health effects in response to short-term exposure to UFPs. While there are a limited number of studies that have examined the association between UFPs and cardiovascular morbidity, there is a larger body of evidence from studies that exposed subjects to fresh DE, which is typically dominated by UFPs. These studies have consistently demonstrated changes in vasomotor function following exposure to atmospheres containing relatively high concentrations of particles (Section 6.2.4.2). Markers of systemic oxidative stress have also been observed to increase after exposure to various particle types that are predominantly in the UFP size range. In addition, alterations in HRV parameters have been observed in response to controlled human exposure to UF CAPs, with inconsistent evidence for changes in markers of blood coagulation following exposure to UF CAPs

and DE (Sections 6.2.1.2 and 6.2.8.2). A few toxicological studies have also found consistent changes in vasomotor function, which provides coherence with the effects demonstrated in the controlled human exposure studies (Section 6.2.4.3). Additional UFP-induced effects observed in toxicological studies include alterations in HRV, with less consistent effects observed for systemic inflammation and blood coagulation. Only a few epidemiologic studies have examined the effect of UFPs on cardiovascular morbidity and collectively they found inconsistent evidence for an association between UFPs and CVD hospital admissions, but some positive associations for subclinical cardiovascular measures (i.e., arrhythmias and supraventricular beats) (Section 6.2.2.1). These studies were conducted in the U.S. and Europe in areas with mean particle number concentration ranging from ~8,500 to 36,000 particles/cm³. However, UFP number concentrations are highly variable (i.e., concentrations drop off quickly from the road compared to accumulation mode particles), and therefore, more subject to exposure error than accumulation mode particles. In conclusion, the evidence from the studies evaluated **is suggestive of a causal relationship between short-term exposures to UFPs and cardiovascular effects.**

Respiratory Effects

A limited number of epidemiologic studies have examined the potential association between short-term exposure to UFPs and respiratory morbidity. Of the studies evaluated, there is limited, and inconsistent evidence for an association between short-term exposure to UFPs and respiratory symptoms, as well as asthma hospital admissions in locations a median particle number concentration of ~6,200 to a mean of 38,000 particles/cm³ (Section 6.3.10). The spatial and temporal variability of UFPs also affects these associations. Toxicological studies have reported respiratory effects including oxidative, inflammatory, and allergic responses using a number of different UFP types (Section 6.3). Although controlled human exposure studies have not extensively examined the effect of UFPs on respiratory outcomes, a few studies have observed small UFP-induced asymptomatic decreases in pulmonary function. Markers of pulmonary inflammation have been observed to increase in healthy adults following controlled exposures to UFPs, particularly in studies using fresh DE. However, it is important to note that for both controlled human exposure and animal toxicological studies of exposures to fresh DE, the relative contributions of gaseous copollutants to the respiratory effects observed remain unresolved. Thus, the current collective evidence **is suggestive of a causal relationship between short-term exposures to UFPs and respiratory effects.**

2.3.6. Integration of UFP Effects

The controlled human exposure studies evaluated have consistently demonstrated effects on vasomotor function and systemic oxidative stress with additional evidence for alterations in HRV parameters in response to exposure to UF CAPs. The toxicological studies provide coherence for the changes in vasomotor function observed in the controlled human exposure studies. Epidemiologic studies are limited because a national network is not in place to measure UFP in the U.S. UFP concentrations are spatially and temporally variable, which would increase uncertainty and make it difficult to detect associations between health effects and UFPs in epidemiologic studies. In addition, data on the composition of UFPs, the spatial and temporal evolution of UFP size distribution and chemical composition, and potential effects of UFP constituents are sparse. More limited evidence is available regarding the effect of UFPs on respiratory effects. Controlled human exposure studies have not extensively examined the effect of UFPs on respiratory measurements, but a few studies have observed small decrements in pulmonary function and increases in pulmonary inflammation. Additional effects including oxidative, inflammatory, and pro-allergic outcomes have been demonstrated in toxicological studies. Epidemiologic studies have found limited and inconsistent evidence for associations between UFPs and respiratory effects. Overall, a limited number of studies have examined the association between exposure to UFPs and morbidity and mortality. Of the studies evaluated, controlled human exposure and toxicological studies provide the most evidence for UFP-induced cardiovascular and respiratory effects; however, many studies focus on exposure to DE. As a result, it is unclear if the effects observed are due to UFP, larger particles (i.e., PM_{2.5}), or the gaseous components of DE. Additionally, UF CAPs systems

are limited as the atmospheric UFP composition is modified when concentrated, which adds uncertainty to the health effects observed in controlled human exposure studies (Section 1.5.3).

2.4. Policy Relevant Considerations

2.4.1. Potentially Susceptible Populations

Upon evaluating the association between short- and long-term exposure to PM and various health outcomes, studies also attempted to identify populations that are more susceptible to PM (i.e., populations that have a greater likelihood of experiencing health effects related to exposure to an air pollutant (e.g., PM) due to a variety of factors including, but not limited to: genetic or developmental factors, race, gender, life stage, lifestyle (e.g., smoking status and nutrition) or preexisting disease; as well as, population-level factors that can increase an individual's exposure to an air pollutant (e.g., PM) such as socioeconomic status [SES], which encompasses reduced access to health care, low educational attainment, residential location, and other factors). These studies did so by conducting stratified analyses; by examining effects in individuals with an underlying health condition; or by developing animal models that mimic the pathophysiologic conditions associated with an adverse health effect. In addition, numerous studies that focus on only one potentially susceptible population provide supporting evidence on whether a population is susceptible to PM exposure. These studies identified a multitude of factors that could potentially contribute to whether an individual is susceptible to PM (Table 8-2). Although studies have primarily used exposures to PM_{2.5} or PM₁₀, the available evidence suggests that the identified factors may also enhance susceptibility to PM_{10-2.5}.

The examination of susceptible populations to PM exposure allows for the NAAQS to provide an adequate margin of safety for both the general population and for susceptible populations.

During specific periods of life (i.e., childhood and advanced age), individuals may be more susceptible to environmental exposures, which in turn can render them more susceptible to PM-related health effects. An evaluation of age-related health effects suggests that older adults have heightened responses for cardiovascular morbidity with PM exposure. In addition, epidemiologic and toxicological studies provide evidence that indicates children are at an increased risk of PM-related respiratory effects. It should be noted that the health effects observed in children could be initiated by exposures to PM that occurred during key windows of development, such as in utero. Epidemiologic studies that focus on exposures during development have reported inconsistent findings (Section 7.4), but a recent toxicological study suggests that inflammatory responses in pregnant women due to exposure to PM could result in health effects in the developing fetus. Epidemiologic studies have also examined whether additional factors, such as gender, race, or ethnicity modify the association between PM and morbidity and mortality outcomes. Although gender and race do not seem to modify PM risk estimates, limited evidence from two studies conducted in California suggest that Hispanic ethnicity may modify the association between PM and mortality.

Recent epidemiologic and toxicological studies provided evidence that individuals with null alleles or polymorphisms in genes that mediate the antioxidant response to oxidative stress (i.e., GSTM1), regulate enzyme activity (i.e., MTHFR and cSHMT), or regulate levels of procoagulants (i.e., fibrinogen) are more susceptible to PM exposure. However, some studies have shown that polymorphisms in genes (e.g., HFE) can have a protective effect against effects of PM exposure. Additionally, preliminary evidence suggests that PM exposure can impart epigenetic effects (i.e., DNA methylation); however, this requires further investigation.

Collectively, the evidence from epidemiologic and toxicological, and to a lesser extent, controlled human exposure studies, indicate increased susceptibility of individuals with underlying CVDs and respiratory illnesses (i.e., asthma) to PM exposure. Controlled human exposure and toxicological studies provide additional evidence for increased PM-related cardiovascular effects in individuals with underlying respiratory health conditions.

Recently studies have begun to examine the influence of preexisting chronic inflammatory conditions, such as diabetes and obesity, on PM-related health effects. These studies have found some evidence for increased associations for cardiovascular outcomes along with pathophysiologic alterations in markers of inflammation, oxidative stress, and acute phase response. However, more

research is needed to thoroughly examine the affect of PM exposure on obese individuals and to identify the biological pathway(s) that could increase the susceptibility of diabetic and obese individuals to PM. There is also evidence that SES, measured using surrogates such as educational attainment or residential location, modifies the association between PM and morbidity and mortality outcomes. In addition, nutritional status, another surrogate measure of SES, has been shown to have protective effects against PM exposure in individuals that have a higher intake of some vitamins and nutrients. Overall, the epidemiologic, controlled human exposure, and toxicological studies evaluated in this review provide evidence for increased susceptibility for various populations, including children and older adults, people with pre-existing cardiopulmonary diseases, and people with lower SES.

2.4.2. Lag Structure of PM-Morbidity and PM-Mortality Associations

Epidemiologic studies have evaluated the time-frame in which exposure to PM can impart a health effect. PM exposure-response relationships can potentially be influenced by a multitude of factors, such as the underlying susceptibility of an individual (e.g., age, pre-existing diseases), which could increase or decrease the lag times observed.

An attempt has been made to identify whether certain lag periods are more strongly associated with specific health outcomes. The epidemiologic evidence evaluated in the 2004 PM AQCD supported the use of lags of 0-1 days for cardiovascular effects and longer moving averages or distributed lags for respiratory diseases (U.S. EPA, 2004, 056905). However, currently, little consensus exists as to the most appropriate a priori lag times to use when examining morbidity and mortality outcomes. As a result, many investigators have chosen to examine the lag structure of associations between PM concentration and health outcome instead of focusing on a priori lag times. This approach is informative because if effects are cumulative, higher overall risks may exist than would be observed for any given single-day lag.

2.4.2.1. PM-Cardiovascular Morbidity Associations

Most of the studies evaluated that examined the association between cardiovascular hospital admissions and ED visits report associations with short-term PM exposure at lags 0- to 2-days, with more limited evidence for shorter durations (i.e., hours) between exposure and response for some health effects (e.g., onset of MI) (Section 6.2.10). However, these studies have rarely examined alternative lag structures. Controlled human exposure and toxicological studies provide biological plausibility for the health effects observed in the epidemiologic studies at immediate or concurrent day lags. Although the majority of the evidence supports shorter lag times for cardiovascular health effects, a recent study has provided preliminary evidence suggesting that longer lag times (i.e., 14- day distributed lag model) may be plausible for non-ischemic cardiovascular conditions

(Section 6.2.10). Panel studies of short-term exposure to PM and cardiovascular endpoints have also examined the time frame from exposure to health effect using a wide range of lag times. Studies of ECG changes indicating ischemia show effects at lags from several hours to 2 days, while lag times ranging from hours to several week moving averages have been observed in studies of arrhythmias, vasomotor function and blood markers of inflammation, coagulation and oxidative stress

(Section 6.2). The longer lags observed in these panel studies may be explained if the effects of PM are cumulative. Although few studies of cumulative effects have been conducted, toxicological studies have demonstrated PM-dependent progression of atherosclerosis. **It should be noted that PM exposure could also lead to an acute event (e.g., infarction or stroke) in individuals with atherosclerosis that may have progressed in response to cumulative PM exposure. Therefore, effects have been observed at a range of lag periods from a few hours to several days with no clear evidence for any lag period having stronger associations than another.**

2.4.2.2. PM-Respiratory Morbidity Associations

Generally, recent studies of respiratory hospital admissions that evaluate multiple lags, have found effect sizes to be larger when using longer moving averages or distributed lag models. For example, when examining hospital admissions for all respiratory diseases among older adults, the strongest associations were observed when using PM concentrations 2 days prior to the hospital

admission (Section 6.3.8). Longer lag periods were also found to be most strongly associated with asthma hospital admissions and ED visits in children (3-5 days) with some evidence for more immediate effects in older adults (lags of 0 and 1 day), but these observations were not consistent across studies (Section 6.3.8). These variable results could be due to the biological complexity of asthma, which inhibits the identification of a specific lag period. The longer lag times identified in the epidemiologic studies evaluated are biologically plausible considering that PM effects on allergic sensitization and lung immune defenses have been observed in controlled human exposure and toxicological studies. These effects could lead to respiratory illnesses over a longer time course (e.g., within several days respiratory infection may become evident, resulting in respiratory symptoms or a hospital admission). However, inflammatory responses, which contribute to some forms of asthma, may result in symptoms requiring medical care within a shorter time frame (e.g., 0-1 days).

2.4.2.3. PM-Mortality Associations

Epidemiologic studies that focused on the association between short-term PM exposure and mortality (i.e., all-cause, cardiovascular, and respiratory) mostly examined a priori lag structures of either 1 or 0-1 days. Although mortality studies do not often examine alternative lag structures, the selection of the aforementioned a priori lag days has been confirmed in additional studies, with the strongest PM-mortality associations consistently being observed at lag 1 and 0-1-days (Section 6.5). However, of note is recent evidence for larger effect estimates when using a distributed lag model. Epidemiologic studies that examined the association between long-term exposure to PM and mortality have also attempted to identify the latency period from PM exposure to death (Section 7.6.4). **Results of the lag comparisons from several cohort studies indicate that the effects of changes in exposure on mortality are seen within five years, with the strongest evidence for effects observed within the first two years. Additionally, there is evidence, albeit from one study, that the mortality effect had larger cumulative effects spread over the follow-up year and three preceding years.**

2.4.3. PM Concentration-Response Relationship

An important consideration in characterizing the PM-morbidity and mortality association is whether the concentration-response relationship is linear across the full concentration range that is encountered or if there are concentration ranges where there are departures from linearity (i.e., nonlinearity). In this ISA studies have been identified that attempt to characterize the shape of the concentration-response curve along with possible PM “thresholds” (i.e., levels which PM concentrations must exceed in order to elicit a health response). The epidemiologic studies evaluated that examined the shape of the concentration-response curve and the potential presence of a threshold have focused on cardiovascular hospital admissions and ED visits and mortality associated with short-term exposure to PM₁₀ and mortality associated with long-term exposure to PM_{2.5}.

A limited number of studies have been identified that examined the shape of the PM-cardiovascular hospital admission and ED visit concentration-response relationship. Of these studies, some conducted an exploratory analysis during model selection to determine if a linear curve most adequately represented the concentration-response relationship; whereas, only one study conducted an extensive analysis to examine the shape of the concentration-response curve at different concentrations (Section 6.2.10.10). Overall, the limited evidence from the studies evaluated supports the use of a no-threshold, log-linear model, which is consistent with the observations made in studies that examined the PM-mortality relationship.

Although multiple studies have previously examined the PM-mortality concentration-response relationship and whether a threshold exists, more complex statistical analyses continue to be developed to analyze this association. Using a variety of methods and models, most of the studies evaluated support the use of a no-threshold, log-linear model; however, one study did observe heterogeneity in the shape of the concentration-response curve across cities (Section 6.5). Overall, the studies evaluated further support the use of a no-threshold log-linear model, but additional issues such as the influence of heterogeneity in estimates between cities, and the effect of seasonal and regional differences in PM on the concentration-response relationship still require further investigation.

In addition to examining the concentration-response relationship between short-term exposure to PM and mortality, Schwartz et al. (2008, 156963) conducted an analysis of the shape of the concentration-response relationship associated with long-term exposure to PM. Using a variety of statistical methods, the concentration-response curve was found to be indistinguishable from linear, and, therefore, little evidence was observed to suggest that a threshold exists in the association between long-term exposure to PM_{2.5} and the risk of death (Section 7.6).

2.4.4. PM Sources and Constituents Linked to Health Effects

Recent epidemiologic, toxicological, and controlled human exposure studies have evaluated the health effects associated with ambient PM constituents and sources, using a variety of quantitative methods applied to a broad set of PM constituents, rather than selecting a few constituents a priori (Section 6.6). There is some evidence for trends and patterns that link particular ambient PM constituents or sources with specific health outcomes, but there is insufficient evidence to determine whether these patterns are consistent or robust.

For cardiovascular effects, multiple outcomes have been linked to a PM_{2.5} crustal/soil/road dust source, including cardiovascular mortality and ST-segment changes. Additional studies have reported associations between other sources (i.e., traffic and wood smoke/vegetative burning) and cardiovascular outcomes (i.e., mortality and ED visits). Studies that only examined the effects of individual PM_{2.5} constituents found evidence for an association between EC and cardiovascular hospital admissions and cardiovascular mortality. Many studies have also observed associations between other sources (i.e., salt, secondary SO₄²⁻/long-range transport, other metals) and cardiovascular effects, but at this time, there does not appear to be a consistent trend or pattern of effects for those factors.

There is less consistent evidence for associations between PM sources and respiratory health effects, which may be partially due to the fact that fewer source apportionment studies have been conducted that examined respiratory-related outcomes (e.g., hospital admissions) and measures (e.g., lung function). However, there is some evidence for associations between respiratory ED visits and decrements in lung function with secondary SO₄²⁻ PM_{2.5}. In addition, crustal/soil/road dust and traffic sources of PM have been found to be associated with increased respiratory symptoms in asthmatic children and decreased PEF in asthmatic adults. Inconsistent results were observed in those PM_{2.5} studies that used individual constituents to examine associations with respiratory morbidity and mortality, although Cu, Pb, OC, and Zn were related to respiratory health effects in two or more studies.

A few studies have identified PM_{2.5} sources associated with total mortality. These studies found an association between mortality and the PM_{2.5} sources: secondary SO₄²⁻/long-range transport, traffic, and salt. In addition, studies have evaluated whether the variation in associations between PM_{2.5} and mortality or PM₁₀ and mortality reflects differences in PM_{2.5} constituents. PM₁₀-mortality effect estimates were greater in areas with a higher proportion of Ni in PM_{2.5}, but the overall PM₁₀-mortality association was diminished when New York City was excluded in sensitivity analyses in two of the studies. V was also found to modify PM₁₀-mortality effect estimates. When examining the effect of species-to-PM_{2.5} mass proportion on PM_{2.5}-mortality effect estimates, Ni, but not V, was also found to modify the association.

Overall, the results indicate that many constituents of PM can be linked with differing health effects and the evidence is not yet sufficient to allow differentiation of those constituents or sources that are more closely related to specific health outcomes. These findings are consistent with the conclusions of the 2004 PM AQCD (U.S. EPA, 2004, 056905) (i.e., that a number of source types, including motor vehicle emissions, coal combustion, oil burning, and vegetative burning, are associated with health effects). Although the crustal factor of fine particles was not associated with mortality in the 2004 PM AQCD (U.S. EPA, 2004, 056905), recent studies have suggested that PM (both PM_{2.5} and PM_{10-2.5}) from crustal, soil or road dust sources or PM tracers linked to these sources are associated with cardiovascular effects. In addition, PM_{2.5} secondary SO₄²⁻ has been associated with both cardiovascular and respiratory effects.

2.5. Welfare Effects

This section presents key conclusions and scientific judgments regarding causality for welfare effects of PM as discussed in Chapter 9. The effects of particulate NO_x and SO_x have recently been evaluated in the ISA for Oxides of Nitrogen and Sulfur – Ecological Criteria (U.S. EPA, 2008, 157074). That ISA focused on the effects from deposition of gas- and particle-phase pollutants related to ambient NO_x and SO_x concentrations that can lead to acidification and nutrient enrichment. Thus, emphasis in Chapter 9 is placed on the effects of airborne PM, including NO_x and SO_x, on visibility and climate, and on the effects of deposition of PM constituents other than NO_x and SO_x, primarily metals and carbonaceous compounds. EPA's framework for causality, described in Chapter 1, was applied and the causal determinations are highlighted.

Table 2-5. Summary of causality determination for welfare effects.

<u>Welfare Effects</u>	<u>Causality Determination</u>
<u>Effects on Visibility</u>	<u>Causal</u>
<u>Effects on Climate</u>	<u>Causal</u>
<u>Ecological Effects</u>	<u>Likely to be causal</u>
<u>Effects on Materials</u>	<u>Causal</u>

2.5.1. Summary of Effects on Visibility

Visibility impairment is caused by light scattering and absorption by suspended particles and gases. There is strong and consistent evidence that PM is the overwhelming source of visibility impairment in both urban and remote areas. EC and some crustal minerals are the only commonly occurring airborne particle components that absorb light. All particles scatter light, and generally light scattering by particles is the largest of the four light extinction components (i.e., absorption and scattering by gases and particles). Although a larger particle scatters more light than a similarly shaped smaller particle of the same composition, the light scattered per unit of mass is greatest for particles with diameters from ~0.3-1.0 μm. For studies where detailed data on particle composition by size are available, accurate calculations of light extinction can be made. However, routinely available PM speciation data can be used to make reasonable estimates of light extinction using relatively simple algorithms that multiply the concentrations of each of the major PM species by its dry extinction efficiency and by a water growth term that accounts for particle size change as a function of relative humidity for hygroscopic species (e.g., sulfate, nitrate, and sea salt). This permits the visibility impairment associated with each of the major PM components to be separately approximated from PM speciation monitoring data.

Direct optical measurement of light extinction measured by transmissometer, or by combining the PM light scattering measured by integrating nephelometers with the PM light absorption measured by an aethalometer, offer a number of advantages compared to algorithm estimates of light extinction based on PM composition and relative humidity data. The direct measurements are not subject to the uncertainties associated with assumed scattering and absorption efficiencies used in the PM algorithm approach. The direct measurements have higher time resolution (i.e., minutes to hours), which is more commensurate with visibility effects compared with calculated light extinction using routinely available PM speciation data (i.e., 24-h duration).

Particulate sulfate and nitrate have comparable light extinction efficiencies (haze impacts per unit mass concentration) at any relative humidity value. Their light scattering per unit mass concentration increases with increasing relative humidity, and at sufficiently high humidity values (RH>85%) they are the most efficient particulate species contributing to haze. Particulate sulfate is

the dominant source of regional haze in the eastern U.S. (>50% of the particulate light extinction) and an important contributor to haze elsewhere in the country (>20% of particulate light extinction). Particulate nitrate is a minor component of remote-area regional haze in the non-California western and eastern U.S., but an important contributor in much of California and in the upper Midwestern U.S., especially during winter when it is the dominant contributor to particulate light extinction.

EC and OC have the highest dry extinction efficiencies of the major PM species and are responsible for a large fraction of the haze, especially in the northwestern U.S., though absolute concentrations are as high in the eastern U.S. Smoke plume impacts from large wildfires dominate many of the worst haze periods in the western U.S. Carbonaceous PM is generally the largest component of urban excess PM_{2.5} (i.e., the difference between urban and regional background concentration). Western urban areas have more than twice the average concentrations of carbonaceous PM than remote areas sites in the same region. In eastern urban areas PM_{2.5} is dominated by about equal concentrations of carbonaceous and sulfate components, though the usually high relative humidity in the East causes the hydrated sulfate particles to be responsible for about twice as much of the urban haze as that caused by the carbonaceous PM.

PM_{2.5} crustal material (referred to as fine soil) and PM_{10-2.5} are significant contributors to haze for remote areas sites in the arid southwestern U.S. where they contribute a quarter to a third of the haze, with PM_{10-2.5} usually contributing twice that of fine soil. Coarse mass concentrations are as high in the Central Great Plains as in the deserts though there are no corresponding high concentrations of fine soil as in the Southwest. Also the relative contribution to haze by the high coarse mass in the Great Plains is much smaller because of the generally higher haze values caused by the high concentrations of sulfate and nitrate PM in that region.

Visibility has direct significance to people's enjoyment of daily activities and their overall sense of wellbeing. For example, psychological research has demonstrated that people are emotionally affected by poor VAQ such that their overall sense of wellbeing is diminished. Urban visibility has been examined in two types of studies directly relevant to the NAAQS review process: urban visibility preference studies and urban visibility valuation studies. Both types of studies are designed to evaluate individuals' desire for good VAQ where they live, using different metrics. Urban visibility preference studies examine individuals' preferences by investigating the amount of visibility degradation considered unacceptable, while economic studies examine the value an individual places on improving VAQ by eliciting how much the individual would be willing to pay for different amounts of VAQ improvement.

There are three urban visibility preference studies and two additional pilot studies that have been conducted to date that provide useful information on individuals' preferences for good VAQ in the urban setting. The completed studies were conducted in Denver, Colorado, two cities in British Columbia, Canada, and Phoenix, AZ. The additional studies were conducted in Washington, DC. The range of median preference values for an acceptable amount of visibility degradation from the 4 urban areas was approximately 19-33 dv. Measured in terms of visual range (VR), these median acceptable values were between approximately 59 and 20 km.

The economic importance of urban visibility has been examined by a number of studies designed to quantify the benefits (or willingness to pay) associated with potential improvements in urban visibility. Urban visibility valuation research was described in the 2004 PM AQCD (U.S. EPA, 2004, 056905) and the 2005 PM Staff Paper (U.S. EPA, 2005, 090209). Since the mid-1990s, little new information has become available regarding urban visibility valuation (Section 9.2.4).

Collectively, the evidence is sufficient to conclude that **a causal relationship exists between PM and visibility impairment.**

2.5.2. Summary of Effects on Climate

Aerosols affect climate through direct and indirect effects. The direct effect is primarily realized as planet brightening when seen from space because most aerosols scatter most of the visible spectrum light that reaches them. The Intergovernmental Panel on Climate Change (IPCC) Fourth Assessment Report (AR4) (IPCC, 2007, 092765), hereafter IPCC AR4, reported that the radiative forcing from this direct effect was $-0.5 (\pm 0.4) \text{ W/m}^2$ and identified the level of scientific understanding of this effect as 'Medium-low'. The global mean direct radiative forcing effect from individual components of aerosols was estimated for the first time in the IPCC AR4 where they were reported to be (all in W/m^2 units): $-0.4 (\pm 0.2)$ for sulfate, $-0.05 (\pm 0.05)$ for fossil fuel-derived organic

carbon, $+0.2 (\pm 0.15)$ for fossil fuel-derived black carbon (BC), $+0.03 (\pm 0.12)$ for biomass burning, $-0.1 (\pm 0.1)$ for nitrates, and $-0.1 (\pm 0.2)$ for mineral dust. Global loadings of anthropogenic dust and nitrates remain very troublesome to estimate, making the radiative forcing estimates for these constituents particularly uncertain.

Numerical modeling of aerosol effects on climate has sustained remarkable progress since the time of the 2004 PM AQCD (U.S. EPA, 2004, 056905), PM AQCD, though model solutions still display large heterogeneity in their estimates of the direct radiative forcing effect from anthropogenic aerosols. The clear-sky direct radiative forcing over ocean due to anthropogenic aerosols is estimated from satellite instruments to be on the order of $-1.1 (\pm 0.37)$ W/m^2 while model estimates are -0.6 W/m^2 . The models' low bias over ocean is carried through for the global average: global average direct radiative forcing from anthropogenic aerosols is estimated from measurements to range from -0.9 to -1.9 W/m^2 , larger than the estimate of -0.8 W/m^2 from the models.

Aerosol indirect effects on climate are primarily realized as an increase in cloud brightness (termed the 'first indirect' or Twomey effect), changes in precipitation, and possible changes in cloud lifetime. The IPCC AR4 reported that the radiative forcing from the Twomey effect was -0.7 (range: -1.1 to $+4$) and identified the level of scientific understanding of this effect as "Low" in part owing to the very large unknowns concerning aerosol size distributions and important interactions with clouds. Other indirect effects from aerosols are not considered to be radiative forcing.

Taken together, direct and indirect effects from aerosols increase Earth's shortwave albedo or reflectance thereby reducing the radiative flux reaching the surface from the Sun. This produces net climate cooling from aerosols. The current scientific consensus reported by IPCC AR4 is that the direct and indirect radiative forcing from anthropogenic aerosols computed at the top of the atmosphere, on a global average, is about -1.3 (range: -2.2 to -0.5) W/m^2 . While the overall global average effect of aerosols at the top of the atmosphere and at the surface is negative, absorption and scattering by aerosols within the atmospheric column warms the atmosphere between the Earth's surface and top of the atmosphere. In part, this is owing to differences in the distribution of aerosol type and size within the vertical atmospheric column since aerosol type and size distributions strongly affect the aerosol scattering and reradiation efficiencies at different altitudes and atmospheric temperatures. And, although the magnitude of the overall negative radiative forcing at the top of the atmosphere appears large in comparison to the analogous IPCC AR4 estimate of positive radiative forcing from anthropogenic GHG of about $+2.9 (\pm 0.3)$ W/m^2 , the horizontal, vertical, and temporal distributions and the physical lifetimes of these two very different radiative forcing agents are not similar; therefore, the effects do not simply off-set one another.

Overall, the evidence is sufficient to conclude that a causal relationship exists between PM and effects on climate, including both direct effects on radiative forcing and indirect effects that involve cloud feedbacks that influence precipitation formation and cloud lifetimes.

2.5.3. Summary of Ecological Effects of PM

Ecological effects of PM include direct effects to metabolic processes of plant foliage; contribution to total metal loading resulting in alteration of soil biogeochemistry and microbiology, plant growth and animal growth and reproduction; and contribution to total organics loading resulting in bioaccumulation and biomagnification across trophic levels. These effects were well-characterized in the 2004 PM AQCD (U.S. EPA, 2004, 056905). Thus, the summary below builds upon the conclusions provided in that review. PM deposition comprises a heterogeneous mixture of particles differing in origin, size, and chemical composition. Exposure to a given concentration of PM may, depending on the mix of deposited particles, lead to a variety of phytotoxic responses and ecosystem effects. Moreover, many of the ecological effects of PM are due to the chemical constituents (e.g., metals, organics, and ions) and their contribution to total loading within an ecosystem.

Investigations of the direct effects of PM deposition on foliage have suggested little or no effects on foliar processes, unless deposition levels were higher than is typically found in the ambient environment. However, consistent and coherent evidence of direct effects of PM has been found in heavily polluted areas adjacent to industrial point sources such as limestone quarries, cement kilns, and metal smelters (Sections 9.4.3 and 9.4.5.7). Where toxic responses have been

documented, they generally have been associated with the acidity, trace metal content, surfactant properties, or salinity of the deposited materials.

An important characteristic of fine particles is their ability to affect the flux of solar radiation passing through the atmosphere, which can be considered in both its direct and diffuse components. Foliar interception by canopy elements occurs for both up- and down-welling radiation. Therefore,

the effect of atmospheric PM on atmospheric turbidity influences canopy processes both by radiation attenuation and by changing the efficiency of radiation interception in the canopy through conversion of direct to diffuse radiation. Crop yields can be sensitive to the amount of radiation received, and crop losses have been attributed to increased regional haze in some areas of the world such as China (Section 9.4.4). On the other hand, diffuse radiation is more uniformly distributed throughout the canopy and may increase canopy photosynthetic productivity by distributing radiation to lower leaves. The enrichment in photosynthetically active radiation (PAR) present in diffuse radiation may offset a portion of the effect of an increased atmospheric albedo due to atmospheric particles. Further research is needed to determine the effects of PM alteration of radiative flux on the growth of vegetation in the U.S.

The deposition of PM onto vegetation and soil, depending on its chemical composition, can produce responses within an ecosystem. The ecosystem response to pollutant deposition is a direct function of the level of sensitivity of the ecosystem and its ability to ameliorate resulting change. Many of the most important ecosystem effects of PM deposition occur in the soil. Upon entering the soil environment, PM pollutants can alter ecological processes of energy flow and nutrient cycling, inhibit nutrient uptake, change ecosystem structure, and affect ecosystem biodiversity. The soil environment is one of the most dynamic sites of biological interaction in nature. It is inhabited by microbial communities of bacteria, fungi, and actinomycetes, in addition to plant roots and soil macro-fauna. These organisms are essential participants in the nutrient cycles that make elements available for plant uptake. Changes in the soil environment can be important in determining plant and ultimately ecosystem response to PM inputs.

There is strong and consistent evidence from field and laboratory experiments that metal components of PM alter numerous aspects of ecosystem structure and function. Changes in the soil chemistry, microbial communities and nutrient cycling, can result from the deposition of trace metals. Exposures to trace metals are highly variable, depending on whether deposition is by wet or dry processes. Although metals can cause phytotoxicity at high concentrations, few heavy metals (e.g., Cu, Ni, Zn) have been documented to cause direct phytotoxicity under field conditions. Exposure to coarse particles and elements such as Fe and Mg are more likely to occur via dry deposition, while fine particles, which are more often deposited by wet deposition, are more likely to contain elements such as Ca, Cr, Pb, Ni, and V. Ecosystems immediately downwind of major emissions sources can receive locally heavy deposition inputs. Phytochelatins produced by plants as a response to sublethal concentrations of heavy metals are indicators of metal stress to plants. Increased concentrations of phytochelatins across regions and at greater elevation have been associated with increased amounts of forest injury in the northeastern U.S.

Overall, the ecological evidence is sufficient to conclude that **a causal relationship is likely to exist between deposition of PM and a variety of effects on individual organisms and ecosystems, based on information from the previous review and limited new findings in**

this review. However, in many cases, it is difficult to characterize the nature and magnitude of effects and to quantify relationships between ambient concentrations of PM and ecosystem response due to significant data gaps and uncertainties as well as considerable variability that exists in the components of PM and their various ecological effects.

2.5.4. Summary of Effects on Materials

Building materials (metals, stones, cements, and paints) undergo natural weathering processes from exposure to environmental elements (wind, moisture, temperature fluctuations, sunlight, etc.). Metals form a protective film of oxidized metal (e.g., rust) that slows environmentally induced corrosion. However, the natural process of metal corrosion is enhanced by exposure to anthropogenic pollutants. For example, formation of hygroscopic salts increases the duration of surface wetness and enhances corrosion.

A significant detrimental effect of particle pollution is the soiling of painted surfaces and other building materials. Soiling changes the reflectance of opaque materials and reduces the transmission

of light through transparent materials. Soiling is a degradation process that requires remediation by cleaning or washing, and, depending on the soiled surface, repainting. Particulate deposition can result in increased cleaning frequency of the exposed surface and may reduce the usefulness of the soiled material. Attempts have been made to quantify the pollutant exposure levels at which materials damage and soiling have been perceived. However, to date, insufficient data are available to advance the knowledge regarding perception thresholds with respect to pollutant concentration, particle size, and chemical composition. Nevertheless, the evidence is sufficient to conclude that **a causal relationship exists between PM and effects on materials.**

2.6. Summary of Health Effects and Welfare Effects Causal Determinations

This chapter has provided an overview of the underlying evidence used in making the causal determinations for the health and welfare effects and PM size fractions evaluated. This review builds upon the main conclusions of the last PM AQCD (U.S. EPA, 2004, 056905):

⑩ “A growing body of evidence both from epidemiological and toxicological studies... supports the general conclusion that PM_{2.5} (or one or more PM_{2.5} components), acting alone and/or in combination with gaseous copollutants, are likely causally related to cardiovascular and respiratory mortality and morbidity.” (pg 9-79)

⑩ “A much more limited body of evidence is suggestive of associations between short-term (but not long-term) exposures to ambient coarse-fraction thoracic particles... and various mortality and morbidity effects observed at times in some locations. This suggests that PM_{10-2.5}, or some constituent component(s) of PM_{10-2.5}, may contribute under some circumstances to increased human health risks... with somewhat stronger evidence for... associations with morbidity (especially respiratory) endpoints than for mortality.” (pg 9-79 and 9-80)

⑩ “Impairment of visibility in rural and urban areas is directly related to ambient concentrations of fine particles, as modulated by particle composition, size, and hygroscopic characteristics, and by relative humidity.” (pg 9-99)

⑩ “Available evidence, ranging from satellite to in situ measurements of aerosol effects on incoming solar radiation and cloud properties, is strongly indicative of an important role in climate for aerosols, but this role is still poorly quantified.” (pg 9-111)

The evaluation of the epidemiologic, toxicological, and controlled human exposure studies published since the completion of the 2004 PM AQCD have provided additional evidence for PM-related health effects. Table 2-6 provides an overview of the causal determinations for all PM size fractions and health effects. Causal determinations for PM and welfare effects, including visibility, climate, ecological effects, and materials are included in Table 2-7. Detailed discussions of the scientific evidence and rationale for these causal determinations are provided in the subsequent chapters of this ISA.

Table 2-6. Summary of PM causal determinations by exposure duration and health outcome.

Size Fraction	Exposure	Outcome	Causality Determination
<u>Cardiovascular Effects</u>			<u>Causal</u>
<u>PM_{2.5}</u>	<u>Short-term</u>	<u>Respiratory Effects</u>	<u>Likely to be causal</u>
		<u>Central Nervous System</u>	<u>Inadequate</u>
		<u>Mortality</u>	<u>Causal</u>
	<u>Long-term</u>	<u>Cardiovascular Effects</u>	<u>Causal</u>
		<u>Respiratory Effects</u>	<u>Likely to be Causal</u>
		<u>Mortality</u>	<u>Causal</u>
		<u>Reproductive and Developmental</u>	<u>Suggestive</u>
<u>Cancer, Mutagenicity, Genotoxicity</u>			<u>Suggestive</u>
<u>Cardiovascular Effects</u>			<u>Suggestive</u>
<u>PM_{10-2.5}</u>	<u>Short-term</u>	<u>Respiratory Effects</u>	<u>Suggestive</u>
		<u>Central Nervous System</u>	<u>Inadequate</u>
		<u>Mortality</u>	<u>Suggestive</u>
	<u>Long-term</u>	<u>Cardiovascular Effects</u>	<u>Inadequate</u>
		<u>Respiratory Effects</u>	<u>Inadequate</u>
		<u>Mortality</u>	<u>Inadequate</u>
<u>Reproductive and Developmental</u>			<u>Inadequate</u>
<u>Cancer, Mutagenicity, Genotoxicity</u>			<u>Inadequate</u>
<u>Cardiovascular Effects</u>			<u>Suggestive</u>
<u>UFPs</u>	<u>Short-term</u>	<u>Respiratory Effects</u>	<u>Suggestive</u>
		<u>Central Nervous System</u>	<u>Inadequate</u>
		<u>Mortality</u>	<u>Inadequate Cardiovascular</u>
	<u>Long-term</u>	<u>Effects</u>	<u>Inadequate Respiratory</u>
		<u>Effects</u>	<u>Inadequate Mortality</u>
			<u>Inadequate</u>
	<u>Reproductive and Developmental</u>	<u>Inadequate</u>	
<u>Cancer, Mutagenicity, Genotoxicity</u>			<u>Inadequate</u>

Table 2-7. Summary of PM causal determinations for welfare effects

<u>Welfare Effects</u>	<u>Causality Determination</u>
<u>Effects on Visibility</u>	<u>Causal</u>
<u>Effects on Climate</u>	<u>Causal</u>
<u>Ecological Effects</u>	<u>Likely to be causal</u>
<u>Effects on Materials</u>	<u>Causal</u>

Chapter 2 References

American Heart Association (2009). Cardiovascular disease statistics. Retrieved 17-NOV-09, from <http://www.americanheart.org/presenter.jhtml?identifier=4478>. [198920](#)

ATSDR (2006). A study of ambient air contaminants and asthma in New York City: Part A and B. Agency for Toxic Substances and Disease Registry; Public Health Service; U.S. Department of Health and Human Services. Atlanta, GA. http://permanent.access.gpo.gov/lps88357/ASTHMA_BRONX_FINAL_REPORT.pdf. [090132](#)

Bell ML; Ebisu K; Peng RD; Walker J; Samet JM; Zeger SL; Dominic F (2008). Seasonal and regional short-term effects of fine particles on hospital admissions in 202 U.S. counties, 1999-2005. *Am J Epidemiol*, 168: 1301-1310. [156266](#)

Burnett RT; Cakmak S; Brook JR; Krewski D (1997). The role of particulate size and chemistry in the association between summertime ambient air pollution and hospitalization for cardiorespiratory diseases. *Environ Health Perspect*, 105: 614-620. [084194](#)

Burnett RT; Goldberg MS (2003). Size-fractionated particulate mass and daily mortality in eight Canadian cities. Health Effects Institute. Boston, MA. [042798](#)

Burnett RT; Stieb D; Brook JR; Cakmak S; Dales R; Raizenne M; Vincent R; Dann T (2004). Associations between short-term changes in nitrogen dioxide and mortality in Canadian cities. *Arch Environ Occup Health*, 59: 228-236. [086247](#)

Chen LH; Knutsen SF; Shavlik D; Beeson WL; Petersen F; Ghamsary M; Abbey D (2005). The association between fatal coronary heart disease and ambient particulate air pollution: Are females at greater risk? *Environ Health Perspect*, 113: 1723-1729. [087942](#)

Chen Y; Yang Q; Krewski D; Burnett RT; Shi Y; McGrail KM (2005). The effect of coarse ambient particulate matter on first, second, and overall hospital admissions for respiratory disease among the elderly. *Inhal Toxicol*, 17: 649-655. [087555](#)

Chen Y; Yang Q; Krewski D; Shi Y; Burnett RT; McGrail K (2004). Influence of relatively low level of particulate air pollution on hospitalization for COPD in elderly people. *Inhal Toxicol*, 16: 21-25. [087262](#)

Chimonas MA; Gessner BD (2007). Airborne particulate matter from primarily geologic, non-industrial sources at levels below National Ambient Air Quality Standards is associated with outpatient visits for asthma and quick-relief medication prescriptions among children less than 20 years old enrolled in Medicaid in Anchorage, Alaska. *Environ Res*, 103: 397-404. [093261](#)

Delfino RJ; Murphy-Moulton AM; Burnett RT; Brook JR; Becklake MR (1997). Effects of air pollution on emergency room visits for respiratory illnesses in Montreal, Quebec. *Am J Respir Crit Care Med*, 155: 568-576. [082687](#)

Dockery DW; Damokosh AI; Neas LM; Raizenne M; Spengler JD; Koutrakis P; Ware JH; Speizer FE (1996). Health effects of acid aerosols on North American children: respiratory symptoms and illness. *Environ Health Perspect*, 104: 500-505. [046219](#)

Dockery DW; Luttmann-Gibson H; Rich DQ; Link MS; Mittleman MA; Gold DR; Koutrakis P; Schwartz JD; Verrier RL (2005). Association of air pollution with increased incidence of ventricular tachyarrhythmias recorded by implanted cardioverter defibrillators. *Environ Health Perspect*, 113: 670-674. [078995](#)

Dominici F; Peng RD; Bell ML; Pham L; McDermott A; Zeger SL; Samet JL (2006). Fine particulate air pollution and hospital admission for cardiovascular and respiratory diseases. *JAMA*, 295: 1127-1134. [088398](#)

Eftim SE; Samet JM; Janes H; McDermott A; Dominici F (2008). Fine particulate matter and mortality: a comparison of the six cities and American Cancer Society cohorts with a medicare cohort. *Epidemiology*, 19: 209-216. [099104](#)

Enstrom JE (2005). Fine particulate air pollution and total mortality among elderly Californians, 1973-2002. *Inhal Toxicol*, 17: 803-816. [087356](#)

⑩ Note: Hyperlinks to the reference citations throughout this document will take you to the NCEA HERO database (Health and Environmental Research Online) at <http://epa.gov/hero>. HERO is a database of scientific literature used by U.S. EPA in the process of developing science assessments such as the Integrated Science Assessments (ISA) and the Integrated Risk Information System (IRIS).

- Fairley D (2003). Mortality and air pollution for Santa Clara County, California, 1989-1996, In: Revised analyses of time-series studies of air pollution and health. Special report. Health Effects Institute. Boston, MA. <http://www.healtheffects.org/Pubs/TimeSeries.pdf> . [042850](#)
- Franklin M; Koutrakis P; Schwartz J (2008). The role of particle composition on the association between PM2.5 and mortality. *Epidemiology*, 19: 680-689. [097426](#)
- Franklin M; Zeka A; Schwartz J (2007). Association between PM2.5 and all-cause and specific-cause mortality in 27 US communities. *J Expo Sci Environ Epidemiol*, 17: 279-287. [091257](#)
- Fung KY; Khan S; Krewski D; Chen Y (2006). Association between air pollution and multiple respiratory hospitalizations among the elderly in Vancouver, Canada. *Inhal Toxicol*, 18: 1005-1011. [089789](#)
- Goss CH; Newsom SA; Schildcrout JS; Sheppard L; Kaufman JD (2004). Effect of ambient air pollution on pulmonary exacerbations and lung function in cystic fibrosis. *Am J Respir Crit Care Med*, 169: 816-821. [055624](#)
- IPCC (2007). *Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report (AR4) of the Intergovernmental Panel on Climate Change* . Cambridge, UK and New York, NY: Intergovernmental Panel on Climate Change, Cambridge University Press. [092765](#)
- Ito K (2003). Associations of particulate matter components with daily mortality and morbidity in Detroit, Michigan, In: Revised analyses of time-series studies of air pollution and health. Special report. Health Effects Institute. Boston, MA. R828112. <http://www.healtheffects.org/Pubs/TimeSeries.pdf> . [042856](#)
- Ito K; Thurston GD; Silverman RA (2007). Characterization of PM2.5, gaseous pollutants, and meteorological interactions in the context of time-series health effects models. *J Expo Sci Environ Epidemiol*, 17 Suppl 2: S45-S60. [156594](#)
- Kim JJ; Smorodinsky S; Lipsett M; Singer BC; Hodgson AT; Ostro B (2004). Traffic-related air pollution near busy roads: the East Bay children's Respiratory Health Study. *Am J Respir Crit Care Med*, 170: 520-526. [087383](#)
- Klemm RJ; Lipfert FW; Wyzga RE; Gust C (2004). Daily mortality and air pollution in Atlanta: two years of data from ARIES. *Inhal Toxicol*, 16 Suppl 1: 131-141. [056585](#)
- Klemm RJ; Mason R (2003). Replication of reanalysis of Harvard Six-City mortality study. In HEI Special Report: Revised Analyses of Time-Series Studies of Air Pollution and Health, Part II (pp. 165-172). Boston, MA: Health Effects Institute. [042801](#)
- Krewski D; Jerrett M; Burnett RT; Ma R; Hughes E; Shi Y; Turner MC; Pope AC III; Thurston G; Calle EE; Thun MJ (2009). Extended follow-up and spatial analysis of the American Cancer Society study linking particulate air pollution and mortality. Health Effects Institute. Cambridge, MA. Report Nr. 140. [191193](#)
- Laden F; Schwartz J; Speizer FE; Dockery DW (2006). Reduction in fine particulate air pollution and mortality: extended follow-up of the Harvard Six Cities study. *Am J Respir Crit Care Med*, 173: 667-672. [087605](#)
- Lin M; Chen Y; Burnett RT; Villeneuve PJ; Krewski D (2002). The influence of ambient coarse particulate matter on asthma hospitalization in children: case-crossover and time-series analyses. *Environ Health Perspect*, 110: 575-581. [026067](#)
- Lin M; Stieb DM; Chen Y (2005). Coarse particulate matter and hospitalization for respiratory infections in children younger than 15 years in Toronto: a case-crossover analysis. *Pediatrics*, 116: 235-240. [087828](#)
- Lipfert FW; Baty JD; Miller JP; Wyzga RE (2006). PM2.5 constituents and related air quality variables as predictors of survival in a cohort of U.S. military veterans. *Inhal Toxicol*, 18: 645-657. [088756](#)
- Lipfert FW; Morris SC; Wyzga RE (2000). Daily mortality in the Philadelphia metropolitan area and size-classified particulate matter. *J Air Waste Manag Assoc*, 50: 1501-1513. [004088](#)
- Lisabeth LD; Escobar JD; Dvorchak JT; Sanchez BN; Majersik JJ; Brown DL; Smith MA; Morgenstern LB (2008). Ambient air pollution and risk for ischemic stroke and transient ischemic attack. *Ann Neurol*, 64: 53-59. [155939](#)
- Mar TF; Larson TV; Stier RA; Claiborn C; Koenig JQ (2004). An analysis of the association between respiratory symptoms in subjects with asthma and daily air pollution in Spokane, Washington. *Inhal Toxicol*, 16: 809-815. [057309](#)
- Mar TF; Norris GA; Larson TV; Wilson WE; Koenig JQ (2003). Air pollution and cardiovascular mortality in Phoenix, 1995-1997. Health Effects Institute. Cambridge, MA. [042841](#)

- McConnell R; Berhane K; Gilliland F; Molitor J; Thomas D; Lurmann F; Avol E; Gauderman WJ; Peters JM (2003). Prospective study of air pollution and bronchitic symptoms in children with asthma. *Am J Respir Crit Care Med*, 168: 790-797. [049490](#)
- Metzger KB; Klein M; Flanders WD; Peel JL; Mulholland JA; Langberg JJ; Tolbert PE (2007). Ambient air pollution and cardiac arrhythmias in patients with implantable defibrillators. *Epidemiology*, 18: 585-592. [092856](#)
- Metzger KB; Tolbert PE; Klein M; Peel JL; Flanders WD; Todd KH; Mulholland JA; Ryan PB; Frumkin H (2004). Ambient air pollution and cardiovascular emergency department visits. *Epidemiology*, 15: 46-56. [044222](#)
- Miller KA; Siscovick DS; Sheppard L; Shepherd K; Sullivan JH; Anderson GL; Kaufman JD (2007). Long-term exposure to air pollution and incidence of cardiovascular events in women. *N Engl J Med*, 356: 447-458. [090130](#)
- O'Connor GT; Neas L; Vaughn B; Kattan M; Mitchell H; Crain EF; Evans R 3rd; Gruchalla R; Morgan W; Stout J; Adams GK; Lippmann M (2008). Acute respiratory health effects of air pollution on children with asthma in US inner cities. *J Allergy Clin Immunol*, 121: 1133-1139. [156818](#)
- Ostro BD; Broadwin R; Lipsett MJ (2003). Coarse particles and daily mortality in Coachella Valley, California. *Health Effects Institute*. Boston, MA. [042824](#)
- Peel JL; Tolbert PE; Klein M; Metzger KB; Flanders WD; Knox T; Mulholland JA; Ryan PB; Frumkin H (2005). Ambient air pollution and respiratory emergency department visits. *Epidemiology*, 16: 164-174. [056305](#)
- Peng RD; Chang HH; Bell ML; McDermott A; Zeger SL; Samet JM; Dominici F (2008). Coarse particulate matter air pollution and hospital admissions for cardiovascular and respiratory diseases among Medicare patients. *JAMA*, 299: 2172-2179. [156850](#)
- Peters A; Dockery DW; Muller JE; Mittleman MA (2001). Increased particulate air pollution and the triggering of myocardial infarction. *Circulation*, 103: 2810-2815. [016546](#)
- Pope C; Renlund D; Kfoury A; May H; Horne B (2008). Relation of heart failure hospitalization to exposure to fine particulate air pollution. *Am J Cardiol*, 102: 1230-1234. [191969](#)
- Pope CA III; Muhlestein JB; May HT; Renlund DG; Anderson JL; Horne BD (2006). Ischemic heart disease events triggered by short-term exposure to fine particulate air pollution. *Circulation*, 114: 2443-2448. [091246](#)
- Rabinovitch N; Strand M; Gelfand EW (2006). Particulate levels are associated with early asthma worsening in children with persistent disease. *Am J Respir Crit Care Med*, 173: 1098-1105. [088031](#)
- Rabinovitch N; Zhang LN; Murphy JR; Vedal S; Dutton SJ; Gelfand EW (2004). Effects of wintertime ambient air pollutants on asthma exacerbations in urban minority children with moderate to severe disease. *J Allergy Clin Immunol*, 114: 1131-1137. [096753](#)
- Renwick LC; Brown D; Clouter A; Donaldson K (2004). Increased inflammation and altered macrophage chemotactic responses caused by two ultrafine particle types. *Occup Environ Med*, 61: 442-447. [056067](#)
- Rich DQ; Kim MH; Turner JR; Mittleman MA; Schwartz J; Catalano PJ; Dockery DW (2006). Association of ventricular arrhythmias detected by implantable cardioverter defibrillator and ambient air pollutants in the St Louis, Missouri metropolitan area. *Occup Environ Med*, 63: 591-596. [089814](#)
- Rich DQ; Schwartz J; Mittleman MA; Link M; Luttmann-Gibson H; Catalano PJ; Speizer FE; Dockery DW (2005). Association of short-term ambient air pollution concentrations and ventricular arrhythmias. *Am J Epidemiol*, 161: 1123-1132. [079620](#)
- Schwartz J; Coull B; Laden F; Ryan L (2008). The effect of dose and timing of dose on the association between airborne particles and survival. *Environ Health Perspect*, 116: 64-69. [156963](#)
- Schwartz J; Dockery DW; Neas LM (1996). Is daily mortality associated specifically with fine particles? *J Air Waste Manag Assoc*, 46: 927-939. [077325](#)
- Sheppard L; Levy D; Norris G; Larson TV; Koenig JQ (2003). Effects of ambient air pollution and nonelderly asthma hospital admissions in Seattle, Washington, 1987-1994. *Epidemiology*, 10: 23-30. [042826](#)
- Slaughter JC; Kim E; Sheppard L; Sullivan JH; Larson TV; Claiborn C (2005). Association between particulate matter and emergency room visits, hospital admissions and mortality in Spokane, Washington. *J Expo Sci Environ Epidemiol*, 15: 153-159. [073854](#)

Stieb DM; Beveridge RC; Brook JR; Smith-Doiron M; Burnett RT; Dales RE; Beaulieu S; Judek S; Mamedov A (2000). Air pollution, aeroallergens and cardiorespiratory emergency department visits in Saint John, Canada. *J Expo Sci Environ Epidemiol*, 10: 461-477. [011675](#)

Sullivan J; Sheppard L; Schreuder A; Ishikawa N; Siscovick D; Kaufman J (2005). Relation between short-term fine- particulate matter exposure and onset of myocardial infarction. *Epidemiology*, 16: 41-48. [050854](#)

Symons JM; Wang L; Guallar E; Howell E; Dominici F; Schwab M; Ange BA; Samet J; Ondov J; Harrison D; Geyh A (2006). A case-crossover study of fine particulate matter air pollution and onset of congestive heart failure symptom exacerbation leading to hospitalization. *Am J Epidemiol*, 164: 421-433. [091258](#)

Thurston GD; Ito K; Hayes CG; Bates DV; Lippmann M (1994). Respiratory hospital admissions and summertime haze air pollution in Toronto, Ontario: consideration of the role of acid aerosols. *Environ Res*, 65: 271-290. [043921](#)

Tolbert PE; Klein M; Peel JL; Sarnat SE; Sarnat JA (2007). Multipollutant modeling issues in a study of ambient air quality and emergency department visits in Atlanta. *J Expo Sci Environ Epidemiol*, 17: S29-S35. [090316](#)

U.S. EPA (2004). Air quality criteria for particulate matter. U.S. Environmental Protection Agency. Research Triangle Park, NC. EPA/600/P-99/002aF-bF. [056905](#)

U.S. EPA (2005). Review of the national ambient air quality standards for particulate matter: Policy assessment of scientific and technical information OAQPS staff paper. U.S. Environmental Protection Agency. Washington, DC. EPA/452/R-05-005a. http://www.epa.gov/ttn/naaqs/standards/pm/data/pmstaffpaper_20051221.pdf . [090209](#)

U.S. EPA (2008). Integrated science assessment for oxides of nitrogen and sulfur: Ecological criteria. U.S. Environmental Protection Agency. Research Triangle Park, NC. EPA/600/R-08/082F. [157074](#)

Villeneuve PJ; Burnett RT; Shi Y; Krewski D; Goldberg MS; Hertzman C; Chen Y; Brook J (2003). A time-series study of air pollution, socioeconomic status, and mortality in Vancouver, Canada. *J Expo Sci Environ Epidemiol*, 13: 427-435. [055051](#)

Villeneuve PJ; Chen L; Stieb D; Rowe BH (2006). Associations between outdoor air pollution and emergency department visits for stroke in Edmonton, Canada. *Eur J Epidemiol*, 21: 689-700. [090191](#)

Wilson WE; Mar TF; Koenig JQ (2007). Influence of exposure error and effect modification by socioeconomic status on the association of acute cardiovascular mortality with particulate matter in Phoenix. *J Expo Sci Environ Epidemiol*, 17: S11-S19. [157149](#)

Woodruff TJ; Darrow LA; Parker JD (2008). Air pollution and postneonatal infant mortality in the United States, 1999-2002. *Environ Health Perspect*, 116: 110-115. [098386](#)

Zanobetti A; Schwartz J (2006). Air pollution and emergency admissions in Boston, MA. *J Epidemiol Community Health*, 60: 890-895. [090195](#)

Zanobetti A; Schwartz J (2009). The effect of fine and coarse particulate air pollution on mortality: A national analysis. *Environ Health Perspect*, 117: 1-40. [188462](#)

Zeger S; Dominici F; McDermott A; Samet J (2008). Mortality in the Medicare population and chronic exposure to fine particulate air pollution in urban centers (2000-2005). *Environ Health Perspect*, 116: 1614-1619. [191951](#)

Zhang Z; Whitsel E; Quibrera P; Smith R; Liao D; Anderson G; Prineas R (2009). Ambient fine particulate matter exposure and myocardial ischemia in the Environmental Epidemiology of Arrhythmogenesis in the Women's Health Initiative (EEAWHI) study. *Environ Health Perspect*, 117: 751-756. [191970](#)



Reference Manual on Scientific Evidence: Third Edition

ISBN
978-0-309-21421-6

1038 pages
6 x 9
PAPERBACK (2011)

Committee on the Development of the Third Edition of the Reference Manual on Scientific Evidence; Federal Judicial Center; National Research Council

 Add book to cart

 Find similar titles

 Share this PDF



Visit the National Academies Press online and register for...

- ✓ Instant access to free PDF downloads of titles from the
 - NATIONAL ACADEMY OF SCIENCES
 - NATIONAL ACADEMY OF ENGINEERING
 - INSTITUTE OF MEDICINE
 - NATIONAL RESEARCH COUNCIL
- ✓ 10% off print titles
- ✓ Custom notification of new releases in your field of interest
- ✓ Special offers and discounts

Distribution, posting, or copying of this PDF is strictly prohibited without written permission of the National Academies Press. Unless otherwise indicated, all materials in this PDF are copyrighted by the National Academy of Sciences. Request reprint permission for this book

Appendix B, Chapter on Epidemiology
Reference Manual on Scientific Evidence
(25 pages)

Reference Guide on Epidemiology

**miChael d. Green, d. miChAl freedman, and
leon Gordis**

Michael D. Green, J.D., is Bess & Walter Williams Chair in Law, Wake Forest University School of Law, Winston-Salem, North Carolina.

D. Michal Freedman, J.D., Ph.D., M.P.H., is Epidemiologist, Division of Cancer Epidemiology and Genetics, National Cancer Institute, Bethesda, Maryland.

Leon Gordis, M.D., M.P.H., Dr.P.H., is Professor Emeritus of Epidemiology, Johns Hopkins Bloomberg School of Public Health, and Professor Emeritus of Pediatrics, Johns Hopkins School of Medicine, Baltimore, Maryland.

CoNtEnTs

- I. Introduction, 551
- II. What Different Kinds of Epidemiologic Studies Exist? 555
 - A. Experimental and Observational Studies of Suspected Toxic Agents, 555
 - B. Types of Observational Study Design, 556
 - 1. Cohort studies, 557
 - 2. Case-control studies, 559
 - 3. Cross-sectional studies, 560
 - 4. Ecological studies, 561
 - C. Epidemiologic and Toxicologic Studies, 563
- III. How Should Results of an Epidemiologic Study Be Interpreted? 566
 - A. Relative Risk, 566
 - B. Odds Ratio, 568
 - C. Attributable Risk, 570
 - D. Adjustment for Study Groups That Are Not Comparable, 571
- IV. What Sources of Error Might Have Produced a False Result? 572
 - A. What Statistical Methods Exist to Evaluate the Possibility of Sampling Error? 574
 - 1. False positives and statistical significance, 575
 - 2. False negatives, 581
 - 3. Power, 582

- B. What Biases May Have Contributed to an Erroneous Association? 583
 - 1. Selection bias, 583
 - 2. Information bias, 585
 - 3. Other conceptual problems, 590
- C. Could a Confounding Factor Be Responsible for the Study Result? 591
 - 1. What techniques can be used to prevent or limit confounding? 595
 - 2. What techniques can be used to identify confounding factors? 595
 - 3. What techniques can be used to control for confounding factors? 596
- V. General Causation: Is an Exposure a Cause of the Disease? 597
 - A. Is There a Temporal Relationship? 601
 - B. How Strong Is the Association Between the Exposure and Disease? 602
 - C. Is There a Dose–Response Relationship? 603
 - D. Have the Results Been Replicated? 604
 - E. Is the Association Biologically Plausible (Consistent with Existing Knowledge)? 604
 - F. Have Alternative Explanations Been Considered? 605
 - G. What Is the Effect of Ceasing Exposure? 605
 - H. Does the Association Exhibit Specificity? 605
 - I. Are the Findings Consistent with Other Relevant Knowledge? 606
- VI. What Methods Exist for Combining the Results of Multiple Studies? 606
- VII. What Role Does Epidemiology Play in Proving Specific Causation? 608
- VIII. Acknowledgments, 618
- Glossary of Terms, 619
- References on Epidemiology, 630
- References on Law and Epidemiology, 630

I. Introduction

Epidemiology is the field of public health and medicine that studies the incidence, distribution, and etiology of disease in human populations. The purpose of epidemiology is to better understand disease causation and to prevent disease in groups of individuals. Epidemiology assumes that disease is not distributed randomly in a group of individuals and that identifiable subgroups, including those exposed to certain agents, are at increased risk of contracting particular diseases.¹

Judges and juries are regularly presented with epidemiologic evidence as the basis of an expert's opinion on causation.² In the courtroom, epidemiologic research findings are offered to establish or dispute whether exposure to an agent³

1. Although epidemiologists may conduct studies of beneficial agents that prevent or cure disease or other medical conditions, this reference guide refers exclusively to outcomes as diseases, because they are the relevant outcomes in most judicial proceedings in which epidemiology is involved.

2. Epidemiologic studies have been well received by courts deciding cases involving toxic substances. *See, e.g.,* *Siharath v. Sandoz Pharms. Corp.*, 131 F. Supp. 2d 1347, 1356 (N.D. Ga. 2001) (“The existence of relevant epidemiologic studies can be a significant factor in proving general causation in toxic tort cases. Indeed, epidemiologic studies provide ‘the primary generally accepted methodology for demonstrating a causal relation between a chemical compound and a set of symptoms or disease.’” (quoting *Conde v. Velsicol Chem. Corp.*, 804 F. Supp. 972, 1025–26 (S.D. Ohio 1992))), *aff'd*, 295 F.3d 1194 (11th Cir. 2002); *Berry v. CSX Transp., Inc.*, 709 So. 2d 552, 569 (Fla. Dist. Ct. App. 1998). Well-conducted studies are uniformly admitted. 3 *Modern Scientific Evidence: The Law and Science of Expert Testimony* § 23.1, at 187 (David L. Faigman et al. eds., 2007–08) [hereinafter *Modern Scientific Evidence*]. Since *Daubert v. Merrell Dow Pharmaceuticals*, 509 U.S. 579 (1993), the predominant use of epidemiologic studies is in connection with motions to exclude the testimony of expert witnesses. Cases deciding such motions routinely address epidemiology and its implications for the admissibility of expert testimony on causation. Often it is not the investigator who conducted the study who is serving as an expert witness in a case in which the study bears on causation. *See, e.g.,* *Kennedy v. Collagen Corp.*, 161 F.3d 1226 (9th Cir. 1998) (physician is permitted to testify about causation); *DeLuca v. Merrell Dow Pharms., Inc.*, 911 F.2d 941, 953 (3d Cir. 1990) (a pediatric pharmacologist expert's credentials are sufficient pursuant to Fed. R. Evid. 702 to interpret epidemiologic studies and render an opinion based thereon); *Medalen v. Tiger Drylac U.S.A., Inc.*, 269 F. Supp. 2d 1118, 1129 (D. Minn. 2003) (holding toxicologist could testify to general causation but not specific causation); *Burton v. R.J. Reynolds Tobacco Co.*, 181 F. Supp. 2d 1256, 1267 (D. Kan. 2002) (a vascular surgeon was permitted to testify to general causation); *Landrigan v. Celotex Corp.*, 605 A.2d 1079, 1088 (N.J. 1992) (an epidemiologist was permitted to testify to both general causation and specific causation); *Trach v. Fellin*, 817 A.2d 1102, 1117–18 (Pa. Super. Ct. 2003) (an expert who was a toxicologist and pathologist was permitted to testify to general and specific causation).

3. We use the term “agent” to refer to any substance external to the human body that potentially causes disease or other health effects. Thus, drugs, devices, chemicals, radiation, and minerals (e.g., asbestos) are all agents whose toxicity an epidemiologist might explore. A single agent or a number of independent agents may cause disease, or the combined presence of two or more agents may be necessary for the development of the disease. Epidemiologists also conduct studies of individual characteristics, such as blood pressure and diet, which might pose risks, but those studies are rarely of interest in judicial proceedings. Epidemiologists also may conduct studies of drugs and other pharmaceutical products to assess their efficacy and safety.

caused a harmful effect or disease.⁴ Epidemiologic evidence identifies agents that are associated with an increased risk of disease in groups of individuals, quantifies the amount of excess disease that is associated with an agent, and provides a profile of the type of individual who is likely to contract a disease after being exposed to an agent. Epidemiology focuses on the question of general causation (i.e., is the agent capable of causing disease?) rather than that of specific causation (i.e., did it cause disease in a particular individual?).⁵ For example, in the 1950s, Doll and Hill and others published articles about the increased risk of lung cancer in cigarette smokers. Doll and Hill's studies showed that smokers who smoked 10 to 20 cigarettes a day had a lung cancer mortality rate that was about 10 times higher than that for nonsmokers.⁶ These studies identified an association between smoking cigarettes and death from lung cancer that contributed to the determination that smoking causes lung cancer.

However, it should be emphasized that *an association is not equivalent to causation*.⁷ An association identified in an epidemiologic study may or may not be

4. *E.g.*, *Bonner v. ISP Techs., Inc.*, 259 F.3d 924 (8th Cir. 2001) (a worker exposed to organic solvents allegedly suffered organic brain dysfunction); *Burton v. R.J. Reynolds Tobacco Co.*, 181 F. Supp. 2d 1256 (D. Kan. 2002) (cigarette smoking was alleged to have caused peripheral vascular disease); *In re Bextra & Celebrex Mktg. Sales Practices & Prod. Liab. Litig.*, 524 F. Supp. 2d 1166 (N.D. Cal. 2007) (multidistrict litigation over drugs for arthritic pain that caused heart disease); *Ruff v. Ensign-Bickford Indus., Inc.*, 168 F. Supp. 2d 1271 (D. Utah 2001) (chemicals that escaped from an explosives manufacturing site allegedly caused non-Hodgkin's lymphoma in nearby residents); *Castillo v. E.I. du Pont de Nemours & Co.*, 854 So. 2d 1264 (Fla. 2003) (a child born with a birth defect allegedly resulting from mother's exposure to a fungicide).

5. This terminology and the distinction between general causation and specific causation are widely recognized in court opinions. *See, e.g.*, *Norris v. Baxter Healthcare Corp.*, 397 F.3d 878 (10th Cir. 2005); *In re Hanford Nuclear Reservation Litig.*, 292 F.3d 1124, 1129 (9th Cir. 2002) ("Generic causation" has typically been understood to mean the capacity of a toxic agent . . . to cause the illnesses complained of by plaintiffs. If such capacity is established, 'individual causation' answers whether that toxic agent actually caused a particular plaintiff's illness."); *In re Rezulin Prods. Liab. Litig.*, 369 F. Supp. 2d 398, 402 (S.D.N.Y. 2005); *Soldo v. Sandoz Pharms. Corp.*, 244 F. Supp. 2d 434, 524–25 (W.D. Pa. 2003); *Burton v. R.J. Reynolds Tobacco Co.*, 181 F. Supp. 2d 1256, 1266–67 (D. Kan. 2002). For a discussion of specific causation, see *infra* Section VII.

6. Richard Doll & A. Bradford Hill, *Lung Cancer and Other Causes of Death in Relation to Smoking: A Second Report on the Mortality of British Doctors*, 2 *Brit. Med. J.* 1071 (1956).

7. *See Soldo v. Sandoz Pharms. Corp.*, 244 F. Supp. 2d 434, 461 (W.D. Pa. 2003) (Hill criteria [see *infra* Section V] developed to assess whether an association is causal); *Miller v. Pfizer, Inc.*, 196 F. Supp. 2d 1062, 1079–80 (D. Kan. 2002); *Magistrini v. One Hour Martinizing Dry Cleaning*, 180 F. Supp. 2d 584, 591 (D.N.J. 2002) ("[A]n association is not equivalent to causation." (quoting the second edition of this reference guide)); *Zandi v. Wyeth a/k/a Wyeth, Inc.*, No. 27-CV-06-6744, 2007 WL 3224242, at *11 (D. Minn. Oct. 15, 2007).

Association is more fully discussed *infra* Section III. The term is used to describe the relationship between two events (e.g., exposure to a chemical agent and development of disease) that occur more frequently together than one would expect by chance. Association does not necessarily imply a causal effect. Causation is used to describe the association between two events when one event is a necessary link in a chain of events that results in the effect. Of course, alternative causal chains may exist that do not include the agent but that result in the same effect. For general treatment of causation in tort law

Reference Guide on Epidemiology

causal.⁸ Assessing whether an association is causal requires an understanding of the strengths and weaknesses of the study's design and implementation, as well as a judgment about how the study findings fit with other scientific knowledge. It is important to emphasize that all studies have "flaws" in the sense of limitations that add uncertainty about the proper interpretation of the results.⁹ Some flaws are inevitable given the limits of technology, resources, the ability and willingness of persons to participate in a study, and ethical constraints. In evaluating epidemiologic evidence, the key questions, then, are the extent to which a study's limitations compromise its findings and permit inferences about causation.

A final caveat is that employing the results of group-based studies of risk to make a causal determination for an individual plaintiff is beyond the limits of epidemiology. Nevertheless, a substantial body of legal precedent has developed that addresses the use of epidemiologic evidence to prove causation for an individual litigant through probabilistic means, and the law developed in these cases is discussed later in this reference guide.¹⁰

The following sections of this reference guide address a number of critical issues that arise in considering the admissibility of, and weight to be accorded to, epidemiologic research findings. Over the past several decades, courts frequently have confronted the use of epidemiologic studies as evidence and have recognized their utility in proving causation. As the Third Circuit observed in *DeLuca v. Merrell Dow Pharmaceuticals, Inc.*: "The reliability of expert testimony founded on reasoning from epidemiologic data is generally a fit subject for judicial notice; epidemiology is a well-established branch of science and medicine, and epidemiologic evidence has been accepted in numerous cases."¹¹ Indeed,

and that for factual causation to exist an agent must be a necessary link in a causal chain sufficient for the outcome, see Restatement (Third) of Torts: Liability for Physical Harm § 26 (2010). Epidemiologic methods cannot deductively prove causation; indeed, all empirically based science cannot affirmatively prove a causal relation. See, e.g., Stephan F. Lanes, *The Logic of Causal Inference in Medicine*, in *Causal Inference* 59 (Kenneth J. Rothman ed., 1988). However, epidemiologic evidence can justify an inference that an agent causes a disease. See *infra* Section V.

8. See *infra* Section IV.

9. See *In re Phenylpropanolamine (PPA) Prods. Liab. Litig.*, 289 F. Supp. 2d 1230, 1240 (W.D. Wash. 2003) (quoting this reference guide and criticizing defendant's "ex post facto dissection" of a study); *In re Orthopedic Bone Screw Prods. Liab. Litig.*, MDL No. 1014, 1997 U.S. Dist. LEXIS 6441, at *26-*27 (E.D. Pa. May 5, 1997) (holding that despite potential for several biases in a study that "may . . . render its conclusions inaccurate," the study was sufficiently reliable to be admissible); Joseph L. Gastwirth, *Reference Guide on Survey Research*, 36 *Jurimetrics J.* 181, 185 (1996) (review essay) ("One can always point to a potential flaw in a statistical analysis.").

10. See *infra* Section VII.

11. 911 F.2d 941, 954 (3d Cir. 1990); see also *Norris v. Baxter Healthcare Corp.*, 397 F.3d 878, 882 (10th Cir. 2005) (an extensive body of exonerative epidemiologic evidence must be confronted and the plaintiff must provide scientifically reliable contrary evidence); *In re Meridia Prods. Liab. Litig.*, 328 F. Supp. 2d 791, 800 (N.D. Ohio 2004) ("Epidemiologic studies are the primary generally accepted methodology for demonstrating a causal relation between the chemical compound and a set of symptoms or a disease. . . ." (quoting *Conde v. Velsicol Chem. Corp.*, 804 F. Supp. 972,

much more difficult problems arise for courts when there is a paucity of epidemiologic evidence.¹²

Three basic issues arise when epidemiology is used in legal disputes, and the methodological soundness of a study and its implications for resolution of the question of causation must be assessed:

1. Do the results of an epidemiologic study or studies reveal an association between an agent and disease?
2. Could this association have resulted from limitations of the study (bias, confounding, or sampling error), and, if so, from which?
3. Based on the analysis of limitations in Item 2, above, and on other evidence, how plausible is a causal interpretation of the association?

Section II explains the different kinds of epidemiologic studies, and Section III addresses the meaning of their outcomes. Section IV examines concerns about the methodological validity of a study, including the problem of sampling error.¹³ Section V discusses general causation, considering whether an agent is capable of causing disease. Section VI deals with methods for combining the results of multiple epidemiologic studies and the difficulties entailed in extracting a single global measure of risk from multiple studies. Additional legal questions that arise in most toxic substances cases are whether population-based epidemiologic evidence can be used to infer specific causation, and, if so, how. Section VII addresses specific causation—the matter of whether a specific agent caused the disease in a given plaintiff.

1025–26 (S.D. Ohio 1992)); *Brasher v. Sandoz Pharms. Corp.*, 160 F. Supp. 2d 1291, 1296 (N.D. Ala. 2001) (“Unquestionably, epidemiologic studies provide the best proof of the general association of a particular substance with particular effects, but it is not the only scientific basis on which those effects can be predicted.”).

12. See *infra* note 181.

13. For a more in-depth discussion of the statistical basis of epidemiology, see David H. Kaye & David A. Freedman, Reference Guide on Statistics, Section II.A, in this manual, and two case studies: Joseph Sanders, *The Bendectin Litigation: A Case Study in the Life Cycle of Mass Torts*, 43 *Hastings L.J.* 301 (1992); Devra L. Davis et al., *Assessing the Power and Quality of Epidemiologic Studies of Asbestos-Exposed Populations*, 1 *Toxicological & Indus. Health* 93 (1985). See also References on Epidemiology and References on Law and Epidemiology at the end of this reference guide.

II. What Different Kinds of Epidemiologic Studies Exist?

A. *Experimental and Observational Studies of Suspected Toxic Agents*

To determine whether an agent is related to the risk of developing a certain disease or an adverse health outcome, we might ideally want to conduct an experimental study in which the subjects would be randomly assigned to one of two groups: one group exposed to the agent of interest and the other not exposed. After a period of time, the study participants in both groups would be evaluated for the development of the disease. This type of study, called a randomized trial, clinical trial, or true experiment, is considered the gold standard for determining the relationship of an agent to a health outcome or adverse side effect. Such a study design is often used to evaluate new drugs or medical treatments and is the best way to ensure that any observed difference in outcome between the two groups is likely to be the result of exposure to the drug or medical treatment.

Randomization minimizes the likelihood that there are differences in relevant characteristics between those exposed to the agent and those not exposed. Researchers conducting clinical trials attempt to use study designs that are placebo controlled, which means that the group not receiving the active agent or treatment is given an inactive ingredient that appears similar to the active agent under study. They also use double blinding where possible, which means that neither the participants nor those conducting the study know which group is receiving the agent or treatment and which group is given the placebo. However, ethical and practical constraints limit the use of such experimental methodologies to assess the value of agents that are thought to be beneficial to human beings.¹⁴

When an agent's effects are suspected to be harmful, researchers cannot knowingly expose people to the agent.¹⁵ Instead epidemiologic studies typically

14. Although experimental human studies cannot intentionally expose subjects to toxins, they can provide evidence that a new drug or other beneficial intervention also has adverse effects. *See In re Bextra & Celebrex Mktg. Sales Practices & Prod. Liab. Litig.*, 524 F. Supp. 2d 1166, 1181 (N.D. Cal. 2007) (the court relied on a clinical study of Celebrex that revealed increased cardiovascular risk to conclude that the plaintiff's experts' testimony on causation was admissible); *McDarby v. Merck & Co.*, 949 A.2d 223 (N.J. Super. Ct. App. Div. 2008) (explaining how clinical trials of Vioxx revealed an association with heart disease).

15. Experimental studies in which human beings are exposed to agents known or thought to be toxic are ethically proscribed. *See Glastetter v. Novartis Pharms. Corp.*, 252 F.3d 986, 992 (8th Cir. 2001); *Brasher v. Sandoz Pharms. Corp.*, 160 F. Supp. 2d 1291, 1297 (N.D. Ala. 2001). Experimental studies can be used where the agent under investigation is believed to be beneficial, as is the case in the development and testing of new pharmaceutical drugs. *See, e.g., McDarby v. Merck & Co.*, 949 A.2d 223, 270 (N.J. Super. Ct. App. Div. 2008) (an expert witness relied on a clinical trial of a new drug to find the adjusted risk for the plaintiff); *see also* Gordon H. Guyatt, *Using Randomized Trials in*

“observe”¹⁶ a group of individuals who have been exposed to an agent of interest, such as cigarette smoke or an industrial chemical and compare them with another group of individuals who have not been exposed. Thus, the investigator identifies a group of subjects who have been exposed¹⁷ and compares their rate of disease or death with that of an unexposed group. In contrast to clinical studies in which potential risk factors can be controlled, epidemiologic investigations generally focus on individuals living in the community, for whom characteristics other than the one of interest, such as diet, exercise, exposure to other environmental agents, and genetic background, may distort a study’s results. Because these characteristics cannot be controlled directly by the investigator, the investigator addresses their possible role in the relationship being studied by considering them in the design of the study and in the analysis and interpretation of the study results (see *infra* Section IV).¹⁸ We emphasize that the Achilles’ heel of observational studies is the possibility of differences in the two populations being studied with regard to risk factors other than exposure to the agent.¹⁹ By contrast, experimental studies, in which subjects are randomized, generally avoid this problem.

B. Types of Observational Study Design

Several different types of observational epidemiologic studies can be conducted.²⁰ Study designs may be chosen because of suitability for investigating the question of interest, timing constraints, resource limitations, or other considerations.

Most observational studies collect data about both exposure and health outcome in every individual in the study. The two main types of observational studies are cohort studies and case-control studies. A third type of observational study is a cross-sectional study, although cross-sectional studies are rarely useful in identifying toxic agents.²¹ A final type of observational study, one in which data about

Pharmacoepidemiology, in *Drug Epidemiology and Post-Marketing Surveillance* 59 (Brian L. Strom & Giampaolo Velo eds., 1992). Experimental studies also may be conducted that entail the discontinuation of exposure to a harmful agent, such as studies in which smokers are randomly assigned to a variety of smoking cessation programs or have no cessation.

16. Classifying these studies as observational in contrast to randomized trials can be misleading to those who are unfamiliar with the area, because subjects in a randomized trial are observed as well. Nevertheless, the use of the term “observational studies” to distinguish them from experimental studies is widely employed.

17. The subjects may have voluntarily exposed themselves to the agent of interest, as is the case, for example, for those who smoke cigarettes, or subjects may have been exposed involuntarily or even without knowledge to an agent, such as in the case of employees who are exposed to chemical fumes at work.

18. See David A. Freedman, *Oasis or Mirage?* 21 *Chance* 59, 59–61 (Mar. 2008).

19. Both experimental and observational studies are subject to random error. See *infra* Section IV.A.

20. Other epidemiologic studies collect data about the group as a whole, rather than about each individual in the group. These group studies are discussed *infra* Section II.B.4.

21. See *infra* Section II.B.3.

individuals are not gathered, but rather population data about exposure and disease are used, is an ecological study.²²

The difference between cohort studies and case-control studies is that cohort studies measure and compare the incidence of disease in the exposed and unexposed (“control”) groups, while case-control studies measure and compare the frequency of exposure in the group with the disease (the “cases”) and the group without the disease (the “controls”). In a case-control study, the rates of exposure in the cases and the rates in the controls are compared, and the odds of having the disease when exposed to a suspected agent can be compared with the odds when not exposed. The critical difference between cohort studies and case-control studies is that cohort studies begin with exposed people and unexposed people, while case-control studies begin with individuals who are selected based on whether they have the disease or do not have the disease and their exposure to the agent in question is measured. The goal of both types of studies is to determine if there is an association between exposure to an agent and a disease and the strength (magnitude) of that association.

1. Cohort studies

In cohort studies,²³ researchers define a study population without regard to the participants’ disease status. The cohort may be defined in the present and followed forward into the future (prospectively) or it may be constructed retrospectively as of sometime in the past and followed over historical time toward the present. In either case, the researchers classify the study participants into groups based on whether they were exposed to the agent of interest (see Figure 1).²⁴ In a prospective study, the exposed and unexposed groups are followed for a specified length of time, and the proportions of individuals in each group who develop the disease of interest are compared. In a retrospective study, the researcher will determine the proportion of individuals in the exposed group who developed the disease from available records or evidence and compare that proportion with the proportion of another group that was not exposed.²⁵ Thus, as illustrated in Table 1,

22. For thumbnail sketches on all types of epidemiologic study designs, see Brian L. Strom, *Study Designs Available for Pharmacoepidemiology Studies*, in *Pharmacoepidemiology* 17, 21–26 (Brian L. Strom ed., 4th ed. 2005).

23. Cohort studies also are referred to as prospective studies and followup studies.

24. In some studies, there may be several groups, each with a different magnitude of exposure to the agent being studied. Thus, a study of cigarette smokers might include heavy smokers (>3 packs a day), moderate smokers (1 to 2 packs a day), and light smokers (<1 pack a day). See, e.g., Robert A. Rinsky et al., *Benzene and Leukemia: An Epidemiologic Risk Assessment*, 316 *New Eng. J. Med.* 1044 (1987).

25. Sometimes in retrospective cohort studies the researcher gathers historical data about exposure and disease outcome of a cohort. Harold A. Kahn, *An Introduction to Epidemiologic Methods* 39–41 (1983). Irving Selikoff, in his seminal study of asbestotic disease in insulation workers, included several hundred workers who had died before he began the study. Selikoff was able to obtain information about exposure from union records and information about disease from hospital and autopsy

Figure 1. Design of a cohort study.

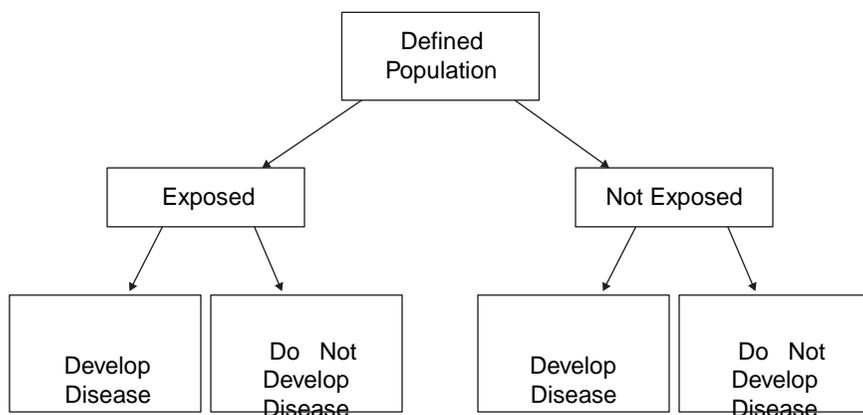


Table 1. Cross-Tabulation of Exposure by Disease Status

	No Disease	Disease	Totals	Incidence Rates of Disease
Not exposed	a	c	$a + c$	$c/(a + c)$
Exposed	b	d	$b + d$	$d/(b + d)$

a researcher would compare the proportion of unexposed individuals with the disease, $c/(a + c)$, with the proportion of exposed individuals with the disease, $d/(b + d)$. If the exposure causes the disease, the researcher would expect a greater proportion of the exposed individuals to develop the disease than the unexposed individuals.²⁶

One advantage of the cohort study design is that the temporal relationship between exposure and disease can often be established more readily than in other study designs, especially a case-control design, discussed below. By tracking people who are initially not affected by the disease, the researcher can determine the time of disease onset and its relation to exposure. This temporal relationship is critical to the question of causation, because exposure must precede disease onset if exposure caused the disease.

As an example, in 1950 a cohort study was begun to determine whether uranium miners exposed to radon were at increased risk for lung cancer as com-

records. Irving J. Selikoff et al., *The Occurrence of Asbestosis Among Insulation Workers in the United States*, 132 Ann. N.Y. Acad. Sci. 139, 143 (1965).

26. Researchers often examine the rate of disease or death in the exposed and control groups. The rate of disease or death entails consideration of the number developing disease within a specified period. All smokers and nonsmokers will, if followed for 100 years, die. Smokers will die at a greater rate than nonsmokers in the earlier years.

pared with nonminers. The study group (also referred to as the exposed cohort) consisted of 3400 white, underground miners. The control group (which need not be the same size as the exposed cohort) comprised white nonminers from the same geographic area. Members of the exposed cohort were examined every 3 years, and the degree of this cohort's exposure to radon was measured from samples taken in the mines. Ongoing testing for radioactivity and periodic medical monitoring of lungs permitted the researchers to examine whether disease was linked to prior work exposure to radiation and allowed them to discern the relationship between exposure to radiation and disease. Exposure to radiation was associated with the development of lung cancer in uranium miners.²⁷

The cohort design is used often in occupational studies such as the one just discussed. Because the design is not experimental, and the investigator has no control over what other exposures a subject in the study may have had, an increased risk of disease among the exposed group may be caused by agents other than the exposure of interest. A cohort study of workers in a certain industry that pays below-average wages might find a higher risk of cancer in those workers. This may be because they work in that industry, or, among other reasons, because low-wage groups are exposed to other harmful agents, such as environmental toxins present in higher concentrations in their neighborhoods. In the study design, the researcher must attempt to identify factors other than the exposure that may be responsible for the increased risk of disease. If data are gathered on other possible etiologic factors, the researcher generally uses statistical methods²⁸ to assess whether a true association exists between working in the industry and cancer. Evaluating whether the association is causal involves additional analysis, as discussed in Section V.

2. Case-control studies

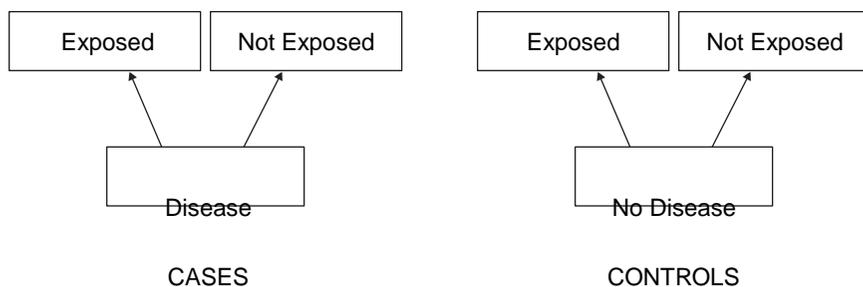
In case-control studies,²⁹ the researcher begins with a group of individuals who have a disease (cases) and then selects a similar group of individuals who do not have the disease (controls). (Ideally, controls should come from the same source population as the cases.) The researcher then compares the groups in terms of past exposures. If a certain exposure is associated with or caused the disease, a higher proportion of past exposure among the cases than among the controls would be expected (see Figure 2).

27. This example is based on a study description in Abraham M. Lilienfeld & David E. Lilienfeld, *Foundations of Epidemiology* 237–39 (2d ed. 1980). The original study is Joseph K. Wagoner et al., *Radiation as the Cause of Lung Cancer Among Uranium Miners*, 273 *New Eng. J. Med.* 181 (1965).

28. See Daniel L. Rubinfield, *Reference Guide on Multiple Regression*, Section II.B, in this manual; David H. Kaye & David A. Freedman, *Reference Guide on Statistics*, Section V.D, in this manual.

29. Case-control studies are also referred to as retrospective studies, because researchers gather historical information about rates of exposure to an agent in the case and control groups.

Figure 2. Design of a case-control study.



Thus, for example, in the late 1960s, doctors in Boston were confronted with an unusual number of young female patients with vaginal adenocarcinoma. Those patients became the “cases” in a case-control study (because they had the disease in question) and were matched with “controls,” who did not have the disease. Controls were selected based on their being born in the same hospitals and at the same time as the cases. The cases and controls were compared for exposure to agents that might be responsible, and researchers found maternal ingestion of DES (diethylstilbestrol) in all but one of the cases but none of the controls.³⁰

An advantage of the case-control study is that it usually can be completed in less time and with less expense than a cohort study. Case-control studies are also particularly useful in the study of rare diseases, because if a cohort study were conducted, an extremely large group would have to be studied in order to observe the development of a sufficient number of cases for analysis.³¹ A number of potential problems with case-control studies are discussed in Section IV.B.

3. Cross-sectional studies

A third type of observational study is a cross-sectional study. In this type of study, individuals are interviewed or examined, and the presence of both the exposure of interest and the disease of interest is determined in each individual at a single point in time. Cross-sectional studies determine the presence (prevalence) of both exposure and disease in the subjects and do not determine the development of disease or risk of disease (incidence). Moreover, because both exposure and disease are determined in an individual at the same point in time, it is not possible to establish the temporal relation between exposure and disease—that is, that the

30. See Arthur L. Herbst et al., *Adenocarcinoma of the Vagina: Association of Maternal Stilbestrol Therapy with Tumor Appearance*, 284 *New Eng. J. Med.* 878 (1971).

31. Thus, for example, to detect a doubling of disease caused by exposure to an agent where the incidence of disease is 1 in 100 in the unexposed population would require sample sizes of 3100 for the exposed and nonexposed groups for a cohort study, but only 177 for the case and control groups in a case-control study. Harold A. Kahn & Christopher T. Sempos, *Statistical Methods in Epidemiology* 66 (1989).

exposure preceded the disease, which would be necessary for drawing any causal inference. Thus, a researcher may use a cross-sectional study to determine the connection between a personal characteristic that does not change over time, such as blood type, and existence of a disease, such as aplastic anemia, by examining individuals and determining their blood types and whether they suffer from aplastic anemia. Cross-sectional studies are infrequently used when the exposure of interest is an environmental toxic agent (current smoking status is a poor measure of an individual's history of smoking), but these studies can provide valuable leads to further directions for research.³²

4. Ecological studies

Up to now, we have discussed studies in which data on both exposure and health outcome are obtained for each individual included in the study.³³ In contrast, studies that collect data only about the group as a whole are called ecological studies.³⁴ In ecological studies, information about individuals is generally not gathered; instead, overall rates of disease or death for different groups are obtained and compared. The objective is to identify some difference between the two groups, such as diet, genetic makeup, or alcohol consumption, that might explain differences in the risk of disease observed in the two groups.³⁵ Such studies may be useful for identifying associations, but they rarely provide definitive causal answers.³⁶ The difficulty is illustrated below with an ecological study of the relationship between dietary fat and cancer.

32. For more information (and references) about cross-sectional studies, see Leon Gordis, *Epidemiology* 195–98 (4th ed. 2009).

33. Some individual studies may be conducted in which all members of a group or community are treated as exposed to an agent of interest (e.g., a contaminated water system) and disease status is determined individually. These studies should be distinguished from ecological studies.

34. In *Cook v. Rockwell International Corp.*, 580 F. Supp. 2d 1071, 1095–96 (D. Colo. 2006), the plaintiffs' expert conducted an ecological study in which he compared the incidence of two cancers among those living in a specified area adjacent to the Rocky Flats Nuclear Weapons Plant with other areas more distant. (The likely explanation for relying on this type of study is the time and expense of a study that gathered information about each individual in the affected area.) The court recognized that ecological studies are less probative than studies in which data are based on individuals but nevertheless held that limitation went to the weight of the study. Plaintiff's expert was permitted to testify to causation, relying on the ecological study he performed.

In *Renaud v. Martin Marietta Corp.*, 749 F. Supp. 1545, 1551 (D. Colo. 1990), *aff'd*, 972 F.2d 304 (10th Cir. 1992), the plaintiffs attempted to rely on an excess incidence of cancers in their neighborhood to prove causation. Unfortunately, the court confused the role of epidemiology in proving causation with the issue of the plaintiffs' exposure to the alleged carcinogen and never addressed the evidentiary value of the plaintiffs' evidence of a disease cluster (i.e., an unusually high incidence of a particular disease in a neighborhood or community). *Id.* at 1554.

35. David E. Lilienfeld & Paul D. Stolley, *Foundations of Epidemiology* 12 (3d ed. 1994).

36. Thus, the emergence of a cluster of adverse events associated with use of heparin, a longtime and widely-prescribed anticoagulant, led to suspicions that some specific lot of heparin was responsible. These concerns led the Centers for Disease Control to conduct a case control study that concluded

If a researcher were interested in determining whether a high dietary fat intake is associated with breast cancer, he or she could compare different countries in terms of their average fat intakes and their average rates of breast cancer. If a country with a high average fat intake also tends to have a high rate of breast cancer, the finding would suggest an association between dietary fat and breast cancer. However, such a finding would be far from conclusive, because it lacks particularized information about an individual's exposure and disease status (i.e., whether an individual with high fat intake is more likely to have breast cancer).³⁷ In addition to the lack of information about an individual's intake of fat, the researcher does not know about the individual's exposures to other agents (or other factors, such as a mother's age at first birth) that may also be responsible for the increased risk of breast cancer. This lack of information about each individual's exposure to an agent and disease status detracts from the usefulness of the study and can lead to an erroneous inference about the relationship between fat intake and breast cancer, a problem known as an ecological fallacy. The fallacy is assuming that, on average, the individuals in the study who have suffered from breast cancer consumed more dietary fat than those who have not suffered from the disease. This assumption may not be true. Nevertheless, the study is useful in that it identifies an area for further research: the fat intake of individuals who have breast cancer as compared with the fat intake of those who do not. Researchers who identify a difference in disease or death in an ecological study may follow up with a study based on gathering data about individuals.

Another epidemiologic approach is to compare disease rates over time and focus on disease rates before and after a point in time when some event of interest took place.³⁸ For example, thalidomide's teratogenicity (capacity to cause birth defects) was discovered after Dr. Widukind Lenz found a dramatic increase in the incidence of limb reduction birth defects in Germany beginning in 1960. Yet, other than with such powerful agents as thalidomide, which increased the incidence of limb reduction defects by several orders of magnitude, these secular-trend studies (also known as time-line studies) are less reliable and less able to

that contaminated heparin manufactured by Baxter was responsible for the outbreak of adverse events. See David B. Blossom et al., *Outbreak of Adverse Event Reactions Associated with Contaminated Heparin*, 359 New Eng. J. Med. 2674 (2008); *In re Heparin Prods. Liab. Litig.* 2011 WL 2971918 (N.D. Ohio July 21, 2011).

37. For a discussion of the data on this question and what they might mean, see David Freedman et al., *Statistics* (4th ed. 2007).

38. In *Wilson v. Merrell Dow Pharmaceuticals, Inc.*, 893 F.2d 1149, 1152–53 (10th Cir. 1990), the defendant introduced evidence showing total sales of Bendectin and the incidence of birth defects during the 1970–1984 period. In 1983, Bendectin was removed from the market, but the rate of birth defects did not change. The Tenth Circuit affirmed the lower court's ruling that the time-line data were admissible and that the defendant's expert witnesses could rely on them in rendering their opinions. Similar evidence was relied on in cases involving cell phones and the drug Parlodel, which was alleged to cause postpartum strokes in women who took the drug to suppress lactation. See *Newman v. Motorola, Inc.*, 218 F. Supp. 2d 769, 778 (D. Md. 2002); *Siharath v. Sandoz Pharms. Corp.*, 131 F. Supp. 2d 1347, 1358 (N.D. Ga. 2001).

detect modest causal effects than the observational studies described above. Other factors that affect the measurement or existence of the disease, such as improved diagnostic techniques and changes in lifestyle or age demographics, may change over time. If those factors can be identified and measured, it may be possible to control for them with statistical methods. Of course, unknown factors cannot be controlled for in these or any other kind of epidemiologic studies.

C. *Epidemiologic and Toxicologic Studies*

In addition to observational epidemiology, toxicology models based on live animal studies (in vivo) may be used to determine toxicity in humans.³⁹ Animal studies have a number of advantages. They can be conducted as true experiments, and researchers control all aspects of the animals' lives. Thus, they can avoid the problem of confounding,⁴⁰ which epidemiology often confronts. Exposure can be carefully controlled and measured. Refusals to participate in a study are not an issue, and loss to followup very often is minimal. Ethical limitations are diminished, and animals can be sacrificed and their tissues examined, which may improve the accuracy of disease assessment. Animal studies often provide useful information about pathological mechanisms and play a complementary role to epidemiology by assisting researchers in framing hypotheses and in developing study designs for epidemiologic studies.

Animal studies have two significant disadvantages, however. First, animal study results must be extrapolated to another species—human beings—and differences in absorption, metabolism, and other factors may result in interspecies variation in responses. For example, one powerful human teratogen, thalidomide, does not cause birth defects in most rodent species.⁴¹ Similarly, some known teratogens in animals are not believed to be human teratogens. In general, it is often difficult to confirm that an agent known to be toxic in animals is safe for human beings.⁴² The second difficulty with inferring human causation from animal studies is that the high doses customarily used in animal studies require consideration of the dose–response relationship and whether a threshold no-effect dose exists.⁴³ Those matters are almost always fraught with considerable, and currently unresolvable, uncertainty.⁴⁴

39. For an in-depth discussion of toxicology, see Bernard D. Goldstein & Mary Sue Henifin, Reference Guide on Toxicology, in this manual.

40. See *infra* Section IV.C.

41. Phillip Knightley et al., *Suffer the Children: The Story of Thalidomide 271–72* (1979).

42. See Ian C.T. Nesbit & Nathan J. Karch, *Chemical Hazards to Human Reproduction 98–106* (1983); Int'l Agency for Research on Cancer (IARC), *Interpretation of Negative Epidemiologic Evidence for Carcinogenicity* (N.J. Wald & Richard Doll eds., 1985) [hereafter IARC].

43. See *infra* Section V.C & note 119.

44. See *Soldo v. Sandoz Pharms. Corp.*, 244 F. Supp. 2d 434, 466 (W.D. Pa. 2003) (quoting this reference guide in the first edition of the Reference Manual); see also *General Elec. Co. v. Joiner*, 522 U.S. 136, 143–45 (1997) (holding that the district court did not abuse its discretion in exclud-

Toxicologists also use in vitro methods, in which human or animal tissue or cells are grown in laboratories and are exposed to certain substances. The problem with this approach is also extrapolation—whether one can generalize the findings from the artificial setting of tissues in laboratories to whole human beings.⁴⁵

Often toxicologic studies are the only or best available evidence of toxicity.⁴⁶ Epidemiologic studies are difficult, time-consuming, expensive, and sometimes, because of limited exposure or the infrequency of disease, virtually impossible to perform.⁴⁷ Consequently, they do not exist for a large array of environmental agents. Where both animal toxicologic and epidemiologic studies are available, no universal rules exist for how to interpret or reconcile them.⁴⁸ Careful assess-

ing expert testimony on causation based on expert's failure to explain how animal studies supported expert's opinion that agent caused disease in humans).

45. For a further discussion of these issues, see Bernard D. Goldstein & Mary Sue Henifin, Reference Guide on Toxicology, Section III.A, in this manual.

46. IARC, a well-regarded international public health agency, evaluates the human carcinogenicity of various agents. In doing so, IARC obtains all of the relevant evidence, including animal studies as well as any human studies. On the basis of a synthesis and evaluation of that evidence, IARC publishes a monograph containing that evidence and its analysis of the evidence and provides a categorical assessment of the likelihood the agent is carcinogenic. In a preamble to each of its monographs, IARC explains what each of the categorical assessments means. Solely on the basis of the strength of animal studies, IARC may classify a substance as “probably carcinogenic to humans.” International Agency for Research on Cancer, *Human Papillomaviruses*, 90 Monographs on the Evaluation of Carcinogenic Risks to Humans 9–10 (2007), available at <http://monographs.iarc.fr/ENG/Monographs/vol90/index.php>; see also *Magistrini v. One Hour Martinizing Dry Cleaning*, 180 F. Supp. 2d 584, 600 n.18 (D.N.J. 2002). When IARC monographs are available, they are generally recognized as authoritative. Unfortunately, IARC has conducted evaluations of only a fraction of potentially carcinogenic agents, and many suspected toxic agents cause effects other than cancer.

47. Thus, in a series of cases involving Parlodel, a lactation suppressant for mothers of newborns, efforts to conduct an epidemiologic study of its effect on causing strokes were stymied by the infrequency of such strokes in women of child-bearing age. See, e.g., *Brasher v. Sandoz Pharms. Corp.*, 160 F. Supp. 2d 1291, 1297 (N.D. Ala. 2001). In other cases, a plaintiff's exposure to an overdose of a drug may be unique or nearly so. See *Zuchowicz v. United States*, 140 F.3d 381 (2d Cir. 1998).

48. See IARC, *supra* note 41 (identifying a number of substances and comparing animal toxicology evidence with epidemiologic evidence); Michele Carbone et al., *Modern Criteria to Establish Human Cancer Etiology*, 64 *Cancer Res.* 5518, 5522 (2004) (National Cancer Institute symposium concluding that “There should be no hierarchy [among different types of scientific methods to determine cancer causation]. Epidemiology, animal, tissue culture and molecular pathology should be seen as integrating evidences in the determination of human carcinogenicity.”)

A number of courts have grappled with the role of animal studies in proving causation in a toxic substance case. One line of cases takes a very dim view of their probative value. For example, in *Brock v. Merrell Dow Pharmaceuticals, Inc.*, 874 F.2d 307, 313 (5th Cir. 1989), the court noted the “very limited usefulness of animal studies when confronted with questions of toxicity.” A similar view is reflected in *Richardson v. Richardson-Merrell, Inc.*, 857 F.2d 823, 830 (D.C. Cir. 1988), *Bell v. Swift Adhesives, Inc.*, 804 F. Supp. 1577, 1579–80 (S.D. Ga. 1992), and *Cadarian v. Merrell Dow Pharmaceuticals, Inc.*, 745 F. Supp. 409, 412 (E.D. Mich. 1989).

Other courts have been more amenable to the use of animal toxicology in proving causation. Thus, in *Marder v. G.D. Searle & Co.*, 630 F. Supp. 1087, 1094 (D. Md. 1986), *aff'd sub nom. Wheelahan v. G.D. Searle & Co.*, 814 F.2d 655 (4th Cir. 1987), the court observed: “There is a range of scientific

ment of the methodological validity and power⁴⁹ of the epidemiologic evidence must be undertaken, and the quality of the toxicologic studies and the questions of interspecies extrapolation and dose–response relationship must be considered.⁵⁰

methods for investigating questions of causation—for example, toxicology and animal studies, clinical research, and epidemiology—which all have distinct advantages and disadvantages.” In *Milward v. Acuity Specialty Products Group, Inc.*, 639 F.3d 11, 17-19 (1st Cir. 2011), the court endorsed an expert’s use of a “weight-of-the-evidence” methodology, holding that the district court abused its discretion in ruling inadmissible an expert’s testimony about causation based on that methodology. As a corollary to recognizing weight of the evidence as a valid scientific technique, the court also noted the role of judgment in making an appropriate inference from the evidence. While recognizing the legitimacy of the methodology, the court also acknowledged that, as with any scientific technique, it can be improperly applied. See also *Metabolife Int’l, Inc. v. Wornick*, 264 F.3d 832, 842 (9th Cir. 2001) (holding that the lower court erred in per se dismissing animal studies, which must be examined to determine whether they are appropriate as a basis for causation determination); *In re Heparin Prods. Liab. Litig.* 2011 WL 2971918 (N.D. Ohio July 21, 2011) (holding that animal toxicology in conjunction with other non-epidemiologic evidence can be sufficient to prove causation); *Ruff v. Ensign-Bickford Indus., Inc.*, 168 F. Supp. 2d 1271, 1281 (D. Utah 2001) (affirming animal studies as sufficient basis for opinion on general causation.); cf. *In re Paoli R.R. Yard PCB Litig.*, 916 F.2d 829, 853–54 (3d Cir. 1990) (questioning the exclusion of animal studies by the lower court). The Third Circuit in a subsequent opinion in *Paoli* observed:

[I]n order for animal studies to be admissible to prove causation in humans, there must be good grounds to extrapolate from animals to humans, just as the methodology of the studies must constitute good grounds to reach conclusions about the animals themselves. Thus, the requirement of reliability, or “good grounds,” extends to each step in an expert’s analysis all the way through the step that connects the work of the expert to the particular case.

In re Paoli R.R. Yard PCB Litig., 35 F.3d 717, 743 (3d Cir. 1994); see also *Cavallo v. Star Enter.*, 892 F. Supp. 756, 761–63 (E.D. Va. 1995) (courts must examine each of the steps that lead to an expert’s opinion), *aff’d in part and rev’d in part*, 100 F.3d 1150 (4th Cir. 1996).

One explanation for these conflicting lines of cases may be that when there is a substantial body of epidemiologic evidence that addresses the causal issue, animal toxicology has much less probative value. That was the case, for example, in the Bendectin cases of *Richardson, Brock, and Cadarian*. Where epidemiologic evidence is not available, animal toxicology may be thought to play a more prominent role in resolving a causal dispute. See Michael D. Green, *Expert Witnesses and Sufficiency of Evidence in Toxic Substances Litigation: The Legacy of Agent Orange and Bendectin Litigation*, 86 Nw. U. L. Rev. 643, 680–82 (1992) (arguing that plaintiffs should be required to prove causation by a preponderance of the available evidence); *Turpin v. Merrell Dow Pharms., Inc.*, 959 F.2d 1349, 1359 (6th Cir. 1992); *In re Paoli R.R. Yard PCB Litig.*, No. 86-2229, 1992 U.S. Dist. LEXIS 16287, at *16 (E.D. Pa. 1992). For another explanation of these cases, see Gerald W. Boston, *A Mass-Exposure Model of Toxic Causation: The Control of Scientific Proof and the Regulatory Experience*, 18 Colum. J. Envtl. L. 181 (1993) (arguing that epidemiologic evidence should be required in mass-exposure cases but not in isolated-exposure cases); see also IARC, *supra* note 41; Bernard D. Goldstein & Mary Sue Henifin, Reference Guide on Toxicology, Section I.F, in this manual. The Supreme Court, in *General Electric Co. v. Joiner*, 522 U.S. 136, 144–45 (1997), suggested that there is no categorical rule for toxicologic studies, observing, “[W]hether animal studies can ever be a proper foundation for an expert’s opinion [is] not the issue. . . . The [animal] studies were so dissimilar to the facts presented in this litigation that it was not an abuse of discretion for the District Court to have rejected the experts’ reliance on them.”

49. See *infra* Section IV.A.3.

50. See Ellen F. Heineman & Shelia Hoar Zahm, *The Role of Epidemiology in Hazard Evaluation*, 9 Toxic Substances J. 255, 258–62 (1989).

III. How Should Results of an Epidemiologic Study Be Interpreted?

Epidemiologists are ultimately interested in whether a causal relationship exists between an agent and a disease. However, the first question an epidemiologist addresses is whether an association exists between exposure to the agent and disease. An association between exposure to an agent and disease exists when they occur together more frequently than one would expect by chance.⁵¹ Although a causal relationship is one possible explanation for an observed association between an exposure and a disease, an association does not necessarily mean that there is a cause–effect relationship. Interpreting the meaning of an observed association is discussed below.

This section begins by describing the ways of expressing the existence and strength of an association between exposure and disease. It reviews ways in which an incorrect result can be produced because of the sampling methods used in all observational epidemiologic studies and then examines statistical methods for evaluating whether an association is real or the result of a sampling error.

The strength of an association between exposure and disease can be stated in various ways,⁵² including as a relative risk, an odds ratio, or an attributable risk.⁵³ Each of these measurements of association examines the degree to which the risk of disease increases when individuals are exposed to an agent.

A. Relative Risk

A commonly used approach for expressing the association between an agent and disease is relative risk (“RR”). It is defined as the ratio of the incidence rate (often referred to as incidence) of disease in exposed individuals to the incidence rate in unexposed individuals:

$$RR = \frac{\text{(Incidence rate in the exposed)}}{\text{(Incidence rate in the unexposed)}}$$

51. A negative association implies that the agent has a protective or curative effect. Because the concern in toxic substances litigation is whether an agent caused disease, this reference guide focuses on positive associations.

52. Another outcome measure is a risk difference. A risk difference is the difference between the proportion of disease in those exposed to the agent and the proportion of disease in those who were unexposed. Thus, in the example of relative risk in the text below discussing relative risk, the proportion of disease in those exposed is 40/100 and the proportion of disease in the unexposed is 20/100. The risk difference is 20/100.

53. Numerous courts have employed these measures of the strength of an association. *See, e.g., In re Bextra & Celebrex Mktg. Sales Practices & Prod. Liab. Litig.*, 524 F. Supp. 2d 1166, 1172–74 (N.D. Cal. 2007); *Cook v. Rockwell Int’l Corp.*, 580 F. Supp. 2d 1071, 1095 (D. Colo. 2006) (citing the second edition of this reference guide); *In re W.R. Grace & Co.*, 355 B.R. 462, 482–83 (Bankr. D. Del. 2006).

The incidence rate of disease is defined as the number of cases of disease that develop during a specified period of time divided by the number of persons in the cohort under study.⁵⁴ Thus, the incidence rate expresses the risk that a member of the population will develop the disease within a specified period of time.

For example, a researcher studies 100 individuals who are exposed to an agent and 200 who are not exposed. After 1 year, 40 of the exposed individuals are diagnosed as having a disease, and 20 of the unexposed individuals also are diagnosed as having the disease. The relative risk of contracting the disease is calculated as follows:

- The incidence rate of disease in the exposed individuals is 40 cases per year per 100 persons (40/100), or 0.4.
- The incidence rate of disease in the unexposed individuals is 20 cases per year per 200 persons (20/200), or 0.1.
- The relative risk is calculated as the incidence rate in the exposed group (0.4) divided by the incidence rate in the unexposed group (0.1), or 4.0.

A relative risk of 4.0 indicates that the risk of disease in the exposed group is four times as high as the risk of disease in the unexposed group.⁵⁵

In general, the relative risk can be interpreted as follows:

- If the relative risk equals 1.0, the risk in exposed individuals is the same as the risk in unexposed individuals.⁵⁶ There is no association between exposure to the agent and disease.
- If the relative risk is greater than 1.0, the risk in exposed individuals is greater than the risk in unexposed individuals. There is a positive association between exposure to the agent and the disease, which could be causal.
- If the relative risk is less than 1.0, the risk in exposed individuals is less than the risk in unexposed individuals. There is a negative association, which could reflect a protective or curative effect of the agent on risk of disease. For example, immunizations lower the risk of disease. The results suggest that immunization is associated with a decrease in disease and may have a protective effect on the risk of disease.

Although relative risk is a straightforward concept, care must be taken in interpreting it. Whenever an association is uncovered, further analysis should be

54. Epidemiologists also use the concept of prevalence, which measures the existence of disease in a population at a given point in time, regardless of when the disease developed. Prevalence is expressed as the proportion of the population with the disease at the chosen time. See Gordis, *supra* note 32, at 43–47.

55. See *DeLuca v. Merrell Dow Pharms., Inc.*, 911 F.2d 941, 947 (3d Cir. 1990); *Magistrini v. One Hour Martinizing Dry Cleaning*, 180 F. Supp. 2d 584, 591 (D.N.J. 2002).

56. See *Magistrini*, 180 F. Supp. 2d at 591.

conducted to assess whether the association is real or a result of sampling error, confounding, or bias.⁵⁷ These same sources of error may mask a true association, resulting in a study that erroneously finds no association.

B. Odds Ratio

The odds ratio (“OR”) is similar to a relative risk in that it expresses in quantitative terms the association between exposure to an agent and a disease.⁵⁸ It is a convenient way to estimate the relative risk in a case-control study when the disease under investigation is rare.⁵⁹ The odds ratio approximates the relative risk when the disease is rare.⁶⁰

In a case-control study, the odds ratio is the ratio of the odds that a case (one with the disease) was exposed to the odds that a control (one without the disease) was exposed. In a cohort study, the odds ratio is the ratio of the odds of developing a disease when exposed to a suspected agent to the odds of developing the disease when not exposed.

Consider a case-control study, with results as shown schematically in a 2×2 table (Table 2):

Table 2. Cross-tabulation of cases and controls by exposure status

	Cases (with disease)	Controls (no disease)
Exposed	<i>a</i>	<i>b</i>
Not exposed	<i>c</i>	<i>d</i>

In a case-control study,

$$\text{OR} = \frac{(\text{Odds that a case was exposed})}{(\text{Odds that a control was exposed})}$$

57. See *infra* Sections IV.B–C.

58. A relative risk cannot be calculated for a case-control study, because a case-control study begins by examining a group of persons who already have the disease. That aspect of the study design prevents a researcher from determining the rate at which individuals develop the disease. Without a rate or incidence of disease, a researcher cannot calculate a relative risk.

59. If the disease is not rare, the odds ratio is still valid to determine whether an association exists, but interpretation of its magnitude is less intuitive.

60. See Marcello Pagano & Kimberlee Gauvreau, *Principles of Biostatistics* 354 (2d ed. 2000). For further detail about the odds ratio and its calculation, see Kahn & Sempos, *supra* note 31, at 47–56.

Looking at Table 2, this ratio can be calculated as

$$\frac{(a/c)}{(b/d)}.$$

This works out to ad/bc . Because we are multiplying two diagonal cells in the table and dividing by the product of the other two diagonal cells, the odds ratio is also called the cross-products ratio.

Consider the following hypothetical study: A researcher identifies 100 individuals with a disease who serve as “cases” and 100 people without the disease who serve as “controls” for her case-control study. Forty of the 100 cases were exposed to the agent and 60 were not. Among the control group, 20 people were exposed and 80 were not. The data can be presented in a 2×2 table (Table 3):

Table 3. Case-Control Study Outcome

	Cases (with disease)	Controls (no disease)
Exposed	40	20
Not exposed	60	80

The calculation of the odds ratio would be:

$$\text{OR} = \frac{(40/60)}{(20/80)} = 2.67.$$

If the disease is relatively rare in the general population (about 5% or less), the odds ratio is a good approximation of the relative risk, which means that there is almost a tripling of the disease in those exposed to the agent.⁶¹

61. The odds ratio is usually marginally greater than the relative risk. As the disease in question becomes more common, the difference between the odds ratio and the relative risk grows.

The reason why the odds ratio approximates the relative risk when the incidence of disease is small can be demonstrated by referring to Table 2. The odds ratio, as stated in the text, is ad/bc . The relative risk for such a study would compare the incidence of disease in the exposed group, or $a/(a+b)$, with the incidence of disease in the unexposed group or $c/(c+d)$. The relative risk would be:

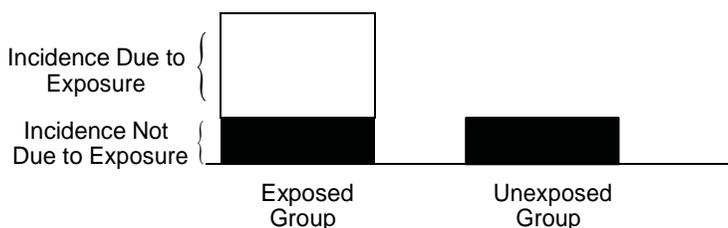
$$\frac{a/(a+b)}{c/(c+d)} = \frac{a/(c+d)}{c/(a+b)}$$

When the incidence of disease is low, a and c will be small in relation to b and d , and the relative risk will then approximate the odds ratio of ad/bc . See Leon Gordis, *Epidemiology* 208–09 (4th ed. 2009).

C. Attributable Risk

A frequently used measurement of risk is the attributable risk (“AR”). The attributable risk represents the amount of disease among exposed individuals that can be attributed to the exposure. It also can be expressed as the proportion of the disease among exposed individuals that is associated with the exposure (also called the “attributable proportion of risk,” the “etiologic fraction,” or the “attributable risk percent”). The attributable risk reflects the maximum proportion of the disease that can be attributed to exposure to an agent and consequently the maximum proportion of disease that could be potentially prevented by blocking the effect of the exposure or by eliminating the exposure.⁶² In other words, if the association is causal, the attributable risk is the proportion of disease in an exposed population that might be caused by the agent and that might be prevented by eliminating exposure to that agent (see Figure 3).⁶³

Figure 3. Risks in exposed and unexposed groups.



To determine the proportion of a disease that is attributable to an exposure, a researcher would need to know the incidence of the disease in the exposed group and the incidence of disease in the unexposed group. The attributable risk is

$$AR = \frac{(\text{incidence in the exposed}) - (\text{incidence in the unexposed})}{\text{incidence in the exposed}}$$

62. Kenneth J. Rothman et al., *Modern Epidemiology* 297 (3d ed. 2008); see also *Landrigan v. Celotex Corp.*, 605 A.2d 1079, 1086 (N.J. 1992) (illustrating that a relative risk of 1.55 conforms to an attributable risk of 35%, that is, $(1.55 - 1.0)/1.55 = .35$, or 35%).

63. Risk is not zero for the control group (those not exposed) when there are other causal chains that cause the disease that do not require exposure to the agent. For example, some birth defects are the result of genetic sources, which do not require the presence of any environmental agent. Also, some degree of risk in the control group may be the result of background exposure to the agent being studied. For example, nonsmokers in a control group may have been exposed to passive cigarette smoke, which is responsible for some cases of lung cancer and other diseases. See also *Ethyl Corp. v. EPA*, 541 F.2d 1, 25 (D.C. Cir. 1976). There are some diseases that do not occur without exposure to an agent; these are known as signature diseases. See *infra* note 177.

The attributable risk can be calculated using the example described in Section III.A. Suppose a researcher studies 100 individuals who are exposed to a substance and 200 who are not exposed. After 1 year, 40 of the exposed individuals are diagnosed as having a disease, and 20 of the unexposed individuals are also diagnosed as having the disease.

- The incidence of disease in the exposed group is 40 persons out of 100 who contract the disease in a year.
- The incidence of disease in the unexposed group is 20 persons out of 200 (or 10 out of 100) who contract the disease in a year.
- The proportion of disease that is attributable to the exposure is 30 persons out of 40, or 75%.

This means that 75% of the disease in the exposed group is attributable to the exposure. We should emphasize here that “attributable” does not necessarily mean “caused by.” Up to this point, we have only addressed associations. Inferring causation from an association is addressed in Section V.

D. Adjustment for Study Groups That Are Not Comparable

Populations often differ in characteristics that relate to disease risk, such as age, sex, and race. Those who live in Florida have a much higher death rate than those who live in Alaska.⁶⁴ Is sunshine dangerous? Perhaps, but the Florida population is much older than the Alaska population, and some adjustment must be made for the differences in age distribution in the two states in order to compare disease or death rates between populations. The technique used to accomplish this is called adjustment, and two types of adjustment are used—direct and indirect. In direct adjustment (e.g., when based on age), overall disease/death rates are calculated for each population as though each had the age distribution of another standard, or reference, population, using the age-specific disease/death rates for each study population. We can then compare these overall rates, called age-adjusted rates, knowing that any difference between these rates cannot be attributed to differences in age, since both age-adjusted rates were generated using the same standard population.

Indirect adjustment is used when the age-specific rates for a study population are not known. In that case, the overall disease/death rate for the standard/reference population is recalculated based on the age distribution of the population of interest using the age-specific rates of the standard population. Then, the actual number of disease cases/deaths in the population of interest can be compared with

64. See Lilienfeld & Stolley, *supra* note 35, at 68–70 (the mortality rate in Florida is approximately three times what it is in Alaska).

the number in the reference population that would be expected if the reference population had the age distribution of the population of interest.

This ratio is called the standardized mortality ratio (SMR). When the outcome of interest is disease rather than death, it is called the standardized morbidity ratio.⁶⁵ If the ratio equals 1.0, the observed number of deaths equals the expected number of deaths, and the mortality rate of the population of interest is no different from that of the reference population. If the SMR is greater than 1.0, the population of interest has a higher mortality risk than that of the reference population, and if the SMR is less than 1.0, the population of interest has a lower mortality rate than that of the reference population.

Thus, age adjustment provides a way to compare populations while in effect holding age constant. Adjustment is used not only for comparing mortality rates in different populations but also for comparing rates in different groups of subjects selected for study in epidemiologic investigations. Although this discussion has focused on adjusting for age, it is also possible to adjust for any number of other variables, such as gender, race, occupation, and socioeconomic status. It is also possible to adjust for several factors simultaneously.⁶⁶

IV. What Sources of Error Might Have Produced a False Result?

Incorrect study results occur in a variety of ways. A study may find a positive association (relative risk greater than 1.0) when there is no true association. Or a study may erroneously result in finding that there is no association when in reality there is. A study may also find an association when one truly exists, but the association found may be greater or less than the real association.

Three general categories of phenomena can result in an association found in a study to be erroneous: chance, bias, and confounding. Before any inferences about causation are drawn from a study, the possibility of these phenomena must be examined.⁶⁷

65. See *Taylor v. Airco, Inc.*, 494 F. Supp. 2d 21, 25 n.4 (D. Mass. 2007) (explaining SMR and its relationship with relative risk). For an example of adjustment used to calculate an SMR for workers exposed to benzene, see Robert A. Rinsky et al., *Benzene and Leukemia: An Epidemiologic Risk Assessment*, 316 New Eng. J. Med. 1044 (1987).

66. For further elaboration on adjustment, see Gordis, *supra* note 32, at 73–78; Philip Cole, *Causality in Epidemiology, Health Policy, and Law*, 27 *Env'tl. L. Rep.* 10,279, 10,281 (1997).

67. See Cole, *supra* note 65, at 10,285. In *DeLuca v. Merrell Dow Pharmaceuticals, Inc.*, 911 F.2d 941, 955 (3d Cir. 1990), the court recognized and discussed random sampling error. It then went on to refer to other errors (e.g., systematic bias) that create as much or more error in the outcome of a study. For a similar description of error in study procedure and random sampling, see David H. Kaye & David A. Freedman, Reference Guide on Statistics, Section IV, in this manual.

The findings of a study may be the result of chance (or random error). In designing a study, the size of the sample can be increased to reduce (but not eliminate) the likelihood of random error. Once a study has been completed, statistical methods (discussed in Section IV.A) permit an assessment of the extent to which the results of a study may be due to random error.

The two main techniques for assessing random error are statistical significance and confidence intervals. A study that is statistically significant has results that are unlikely to be the result of random error, although any criterion for “significance” is somewhat arbitrary. A confidence interval provides both the relative risk (or other risk measure) found in the study and a range (interval) within which the risk likely would fall if the study were repeated numerous times. These two techniques (which are closely related) are explained in Section IV.A.

We should emphasize a matter that those unfamiliar with statistical methodology frequently find confusing: That a study’s results are statistically significant says nothing about the importance of the magnitude of any association (i.e., the relative risk or odds ratio) found in a study or about the biological or clinical importance of the finding.⁶⁸ “Significant,” as used with the adjective “statistically,” does not mean important. A study may find a statistically significant relationship that is quite modest—perhaps it increases the risk only by 5%, which is equivalent to a relative risk of 1.05.⁶⁹ An association may be quite large—the exposed cohort might be 10 times more likely to develop disease than the control group—but the association is not statistically significant because of the potential for random error given a small sample size. In short, *statistical significance is not about the size of the risk found in a study*.

Bias (or systematic error) also can produce error in the outcome of a study. Epidemiologists attempt to minimize bias through their study design, including data collection protocols. Study designs are developed before they begin gathering data. However, even the best designed and conducted studies have biases, which may be subtle. Consequently, after data collection is completed, analytical tools are often used to evaluate potential sources of bias. Sometimes, after bias is identified, the epidemiologist can determine whether the bias would tend to inflate or dilute any association that may exist. Identification of the bias may permit the

68. See Modern Scientific Evidence, *supra* note 2, § 6.36 at 358 (“Statisticians distinguish between ‘statistical’ and ‘practical’ significance. . . .”); Cole, *supra* note 65, at 10,282. Understandably, some courts have been confused about the relationship between statistical significance and the magnitude of the association. See Hyman & Armstrong, P.S.C. v. Gunderson, 279 S.W.3d 93, 102 (Ky. 2008) (describing a small increased risk as being considered statistically insignificant and a somewhat larger risk as being considered statistically significant.); *In re Pfizer Inc. Sec. Litig.*, 584 F. Supp. 2d 621, 634–35 (S.D.N.Y. 2008) (confusing the magnitude of the effect with whether the effect was statistically significant); *In re Joint E. & S. Dist. Asbestos Litig.*, 827 F. Supp. 1014, 1041 (S.D.N.Y. 1993) (concluding that any relative risk less than 1.50 is statistically insignificant), *rev’d on other grounds*, 52 F.3d 1124 (2d Cir. 1995).

69. In general, small effects that are statistically significant require larger sample sizes. When effects are larger, generally fewer subjects are required to produce statistically significant findings.

epidemiologist to make an assessment of whether the study's conclusions are valid. Epidemiologists may reanalyze a study's data to correct for a bias identified in a completed study or to validate the analytical methods used.⁷⁰ Common biases and how they may produce invalid results are described in Section IV.B.

Finally, a study may reach incorrect conclusions about causation because, although the agent and disease are associated, the agent is not a true causal factor. Rather, the agent may be associated with another agent that is the true causal factor, and this latter factor confounds the relationship being examined in the study. Confounding is explained in Section IV.C.

A. *What Statistical Methods Exist to Evaluate the Possibility of Sampling Error?*⁷¹

Before detailing the statistical methods used to assess random error (which we use as synonymous with sampling error), two concepts are explained that are central to epidemiology and statistical analysis. Understanding these concepts should facilitate comprehension of the statistical methods.

Epidemiologists often refer to the true association (also called "real association"), which is the association that really exists between an agent and a disease and that might be found by a perfect (but nonexistent) study. The true association is a concept that is used in evaluating the results of a given study even though its value is unknown. By contrast, a study's outcome will produce an observed association, which is known.

Formal procedures for statistical testing begin with the null hypothesis, which posits that there is no true association (i.e., a relative risk of 1.0) between the agent and disease under study. Data are gathered and analyzed to see whether they disprove⁷² the null hypothesis. The data are subjected to statistical testing to assess the plausibility that any association found is a result of random error or whether it supports rejection of the null hypothesis. The use of the null hypothesis for this testing should not be understood as the a priori belief of the investigator. When epidemiologists investigate an agent, it is usually because they hypothesize that the agent is a cause of some outcome. Nevertheless, epidemiologists prepare their

70. E.g., Richard A. Kronmal et al., *The Intrauterine Device and Pelvic Inflammatory Disease: The Women's Health Study Reanalyzed*, 44 J. Clin. Epidemiol. 109 (1991) (a reanalysis of a study that found an association between the use of IUDs and pelvic inflammatory disease concluded that IUDs do not increase the risk of pelvic inflammatory disease).

71. For a bibliography on the role of statistical significance in legal proceedings, see Sanders, *supra* note 13, at 329 n.138.

72. See, e.g., *Daubert v. Merrell Dow Pharms., Inc.*, 509 U.S. 579, 593 (1993) (scientific methodology involves generating and testing hypotheses).

study designs and test the plausibility that any association found in a study was the result of random error by using the null hypothesis.⁷³

1. False positives and statistical significance

When a study results in a positive association (i.e., a relative risk greater than 1.0), epidemiologists try to determine whether that outcome represents a true association or is the result of random error.⁷⁴ Random error is illustrated by a fair coin (i.e., not modified to produce more heads than tails [or vice versa]). On average, for example, we would expect that coin tosses would yield half heads and half tails. But sometimes, a set of coin tosses might yield an unusual result, for example, six heads out of six tosses,⁷⁵ an occurrence that would result, purely by chance, in less than 2% of a series of six tosses. In the world of epidemiology, sometimes the study findings, merely by chance, do not reflect the true relationships between an agent and outcome. Any single study—even a clinical trial—is in some ways analogous to a set of coin tosses, being subject to the play of chance. Thus, for example, even though the true relative risk (in the total population) is 1.0, an epidemiologic study of a particular study population may find a relative risk greater than (or less

73. See *DeLuca v. Merrell Dow Pharms., Inc.*, 911 F.2d 941, 945 (3d Cir. 1990); *United States v. Philip Morris USA, Inc.*, 449 F. Supp. 2d 1, 706 n.29 (D.D.C. 2006); Stephen E. Fienberg et al., *Understanding and Evaluating Statistical Evidence in Litigation*, 36 *Jurimetrics J.* 1, 21–24 (1995).

74. Hypothesis testing is one of the most counterintuitive techniques in statistics. Given a set of epidemiologic data, one wants to ask the straightforward, obvious question: What is the probability that the difference between two samples reflects a real difference between the populations from which they were taken? Unfortunately, there is no way to answer this question directly or to calculate the probability. Instead, statisticians—and epidemiologists—address a related but very different question: If there really is no difference between the populations, how probable is it that one would find a difference at least as large as the observed difference between the samples? See *Modern Scientific Evidence*, *supra* note 2, § 6:36, at 359 (“it is easy to mistake the *p*-value for the probability that there is no difference”); *Expert Evidence: A Practitioner’s Guide to Law, Science, and the FJC Manual 91* (Bert Black & Patrick W. Lee eds., 1997). Thus, the *p*-value for a given study does not provide a rate of error or even a probability of error for an epidemiologic study. In *Daubert v. Merrell Dow Pharmaceuticals, Inc.*, 509 U.S. 579, 593 (1993), the Court stated that “the known or potential rate of error” should ordinarily be considered in assessing scientific reliability. Epidemiology, however, unlike some other methodologies—fingerprint identification, for example—does not permit an assessment of its accuracy by testing with a known reference standard. A *p*-value provides information only about the plausibility of random error given the study result, but the true relationship between agent and outcome remains unknown. Moreover, a *p*-value provides no information about whether other sources of error—bias and confounding—exist and, if so, their magnitude. In short, for epidemiology, there is no way to determine a rate of error. See *Kumho Tire Co. v. Carmichael*, 526 U.S. 137, 151 (1999) (recognizing that for different scientific and technical inquiries, different considerations will be appropriate for assessing reliability); *Cook v. Rockwell Int’l Corp.*, 580 F. Supp. 2d 1071, 1100 (D. Colo. 2006) (“Defendants have not argued or presented evidence that . . . a method by which an overall ‘rate of error’ can be calculated for an epidemiologic study.”)

75. *DeLuca*, 911 F.2d at 946–47.

than) 1.0 because of random error or chance.⁷⁶ An erroneous conclusion that the null hypothesis is false (i.e., a conclusion that there is a difference in risk when no difference actually exists) owing to random error is called a false-positive error (also Type I error or alpha error).

Common sense leads one to believe that a large enough sample of individuals must be studied if the study is to identify a relationship between exposure to an agent and disease that truly exists. Common sense also suggests that by enlarging the sample size (the size of the study group), researchers can form a more accurate conclusion and reduce the chance of random error in their results. Both statements are correct and can be illustrated by a test to determine if a coin is fair. A test in which a fair coin is tossed 1000 times is more likely to produce close to 50% heads than a test in which the coin is tossed only 10 times. It is far more likely that a test of a fair coin with 10 tosses will come up, for example, with 80% heads than will a test with 1000 tosses. With large numbers, the outcome of the test is less likely to be influenced by random error, and the researcher would have greater confidence in the inferences drawn from the data.⁷⁷

One means for evaluating the possibility that an observed association could have occurred as a result of random error is by calculating a *p*-value.⁷⁸ A *p*-value represents the probability that an observed positive association could result from random error even if no association were in fact present. Thus, a *p*-value of .1 means that there is a 10% chance that values at least as large as the observed relative risk could have occurred by random error, with no association actually present in the population.⁷⁹

To minimize false positives, epidemiologists use a convention that the *p*-value must fall below some selected level known as alpha or significance level for the results of the study to be statistically significant.⁸⁰ Thus, an outcome is statistically significant when the observed *p*-value for the study falls below the preselected

76. See *Magistrini v. One Hour Martinizing Dry Cleaning*, 180 F. Supp. 2d 584, 592 (D.N.J. 2002) (citing the second edition of this reference guide).

77. This explanation of numerical stability was drawn from Brief for Professor Alvan R. Feinstein as Amicus Curiae Supporting Respondents at 12–13, *Daubert v. Merrell Dow Pharms., Inc.*, 509 U.S. 579 (1993) (No. 92-102). See also *Allen v. United States*, 588 F. Supp. 247, 417–18 (D. Utah 1984), *rev'd on other grounds*, 816 F.2d 1417 (10th Cir. 1987). The *Allen* court observed that although “[s]mall communities or groups of people are deemed ‘statistically unstable’” and “data from small populations must be handled with care [, it] does not mean that [the data] cannot provide substantial evidence in aid of our effort to describe and understand events.”

78. See also David H. Kaye & David A. Freedman, Reference Guide on Statistics, Section IV.B, in this manual (the *p*-value reflects the implausibility of the null hypothesis).

79. Technically, a *p*-value of .1 means that if in fact there is no association, 10% of all similar studies would be expected to yield an association the same as, or greater than, the one found in the study due to random error.

80. *Cook v. Rockwell Int'l Corp.*, 580 F. Supp. 2d 1071, 1100–01 (D. Colo. 2006) (discussing *p*-values and their relationship with statistical significance); *Allen*, 588 F. Supp. at 416–17 (discussing statistical significance and selection of a level of alpha); see also Sanders, *supra* note 13, at 343–44 (explaining alpha, beta, and their relationship to sample size); *Developments in the Law—Confronting*

significance level. The most common significance level, or alpha, used in science is .05.⁸¹ A .05 value means that the probability is 5% of observing an association at least as large as that found in the study when in truth there is no association.⁸² Although .05 is often the significance level selected, other levels can and have been used.⁸³ Thus, in its study of the effects of second-hand smoke, the U.S.

the New Challenges of Scientific Evidence, 108 Harv. L. Rev. 1481, 1535–36, 1540–46 (1995) [hereafter *Developments in the Law*].

81. A common error made by lawyers, judges, and academics is to equate the level of alpha with the legal burden of proof. Thus, one will often see a statement that using an alpha of .05 for statistical significance imposes a burden of proof on the plaintiff far higher than the civil burden of a preponderance of the evidence (i.e., greater than 50%). See, e.g., *In re Ephedra Prods. Liab. Litig.*, 393 F. Supp. 2d 181, 193 (S.D.N.Y. 2005); *Marmo v. IBP, Inc.*, 360 F. Supp. 2d 1019, 1021 n.2 (D. Neb. 2005) (an expert toxicologist who stated that science requires proof with 95% certainty while expressing his understanding that the legal standard merely required more probable than not). *But see* *Giles v. Wyeth, Inc.*, 500 F. Supp. 2d 1048, 1056–57 (S.D. Ill. 2007) (quoting the second edition of this reference guide).

Comparing a selected *p*-value with the legal burden of proof is mistaken, although the reasons are a bit complex and a full explanation would require more space and detail than is feasible here. Nevertheless, we sketch out a brief explanation: First, alpha does not address the likelihood that a plaintiff's disease was caused by exposure to the agent; the magnitude of the association bears on that question. See *infra* Section VII. Second, significance testing only bears on whether the observed magnitude of association arose as a result of random chance, not on whether the null hypothesis is true. Third, using stringent significance testing to avoid false-positive error comes at a complementary cost of inducing false-negative error. Fourth, using an alpha of .5 would not be equivalent to saying that the probability the association found is real is 50%, and the probability that it is a result of random error is 50%. Statistical methodology does not permit assessments of those probabilities. See Green, *supra* note 47, at 686; Michael D. Green, *Science Is to Law as the Burden of Proof Is to Significance Testing*, 37 *Jurimetrics J.* 205 (1997) (book review); see also David H. Kaye, *Apples and Oranges: Confidence Coefficients and the Burden of Persuasion*, 73 *Cornell L. Rev.* 54, 66 (1987); David H. Kaye & David A. Freedman, *Reference Guide on Statistics*, Section IV.B.2, in this manual; *Turpin v. Merrell Dow Pharms., Inc.*, 959 F.2d 1349, 1357 n.2 (6th Cir. 1992), *cert. denied*, 506 U.S. 826 (1992); cf. *DeLuca*, 911 F.2d at 959 n.24 (“The relationship between confidence levels and the more likely than not standard of proof is a very complex one . . . and in the absence of more education than can be found in this record, we decline to comment further on it.”).

82. This means that if one conducted an examination of a large number of associations in which the true RR equals 1, on average 1 in 20 associations found to be statistically significant at a .05 level would be spurious. When researchers examine many possible associations that might exist in their data—known as data dredging—we should expect that even if there are no true causal relationships, those researchers will find statistically significant associations in 1 of every 20 associations examined. See Rachel Nowak, *Problems in Clinical Trials Go Far Beyond Misconduct*, 264 *Sci.* 1538, 1539 (1994).

83. A significance test can be either one-tailed or two-tailed, depending on the null hypothesis selected by the researcher. Because most investigators of toxic substances are only interested in whether the agent increases the incidence of disease (as distinguished from providing protection from the disease), a one-tailed test is often viewed as appropriate. *In re Phenylpropanolamine (PPA) Prods. Liab. Litig.*, 289 F. Supp. 2d 1230, 1241 (W.D. Wash. 2003) (accepting the propriety of a one-tailed test for statistical significance in a toxic substance case); *United States v. Philip Morris USA, Inc.*, 449 F. Supp. 2d 1, 701 (D.D.C. 2006) (explaining the basis for EPA's decision to use one-tailed test in assessing whether second-hand smoke was a carcinogen). *But see* *Good v. Fluor Daniel Corp.*, 222 F. Supp. 2d 1236, 1243 (E.D. Wash. 2002). For an explanation of the difference

Environmental Protection Agency (EPA) used a .10 standard for significance testing.⁸⁴

There is some controversy among epidemiologists and biostatisticians about the appropriate role of significance testing.⁸⁵ To the strictest significance testers,

between one-tailed and two-tailed tests, see David H. Kaye & David A. Freedman, Reference Guide on Statistics, Section IV.C.2, in this manual.

84. U.S. Environmental Protection Agency, *Respiratory Health Effects of Passive Smoking: Lung Cancer and Other Disorders* (1992); see also *Turpin*, 959 F.2d at 1353–54 n.1 (confidence level frequently set at 95%, although 90% (which corresponds to an alpha of .10) is also used; selection of the value is “somewhat arbitrary”).

85. Similar controversy exists among the courts that have confronted the issue of whether statistically significant studies are required to satisfy the burden of production. The leading case advocating statistically significant studies is *Brock v. Merrell Dow Pharmaceuticals, Inc.*, 874 F.2d 307, 312 (5th Cir. 1989), *amended*, 884 F.2d 167 (5th Cir.), *cert. denied*, 494 U.S. 1046 (1990). Overturning a jury verdict for the plaintiff in a Bendectin case, the court observed that no statistically significant study had been published that found an increased relative risk for birth defects in children whose mothers had taken Bendectin. The court concluded: “[W]e do not wish this case to stand as a bar to future Bendectin cases in the event that new and statistically significant studies emerge which would give a jury a firmer basis on which to determine the issue of causation.” *Brock*, 884 F.2d at 167.

A number of courts have followed the *Brock* decision or have indicated strong support for significance testing as a screening device. See *Good v. Fluor Daniel Corp.*, 222 F. Supp. 2d 1236, 1243 (E.D. Wash. 2002) (“In the absence of a statistically significant difference upon which to opine, Dr. Au’s opinion must be excluded under *Daubert*.”); *Miller v. Pfizer, Inc.*, 196 F. Supp. 2d 1062, 1080 (D. Kan. 2002) (the expert must have statistically significant studies to serve as basis of opinion on causation); *Kelley v. Am. Heyer-Schulte Corp.*, 957 F. Supp. 873, 878 (W.D. Tex. 1997) (the lower end of the confidence interval must be above 1.0—equivalent to requiring that a study be statistically significant—before a study may be relied upon by an expert), *appeal dismissed*, 139 F.3d 899 (5th Cir. 1998); *Renaud v. Martin Marietta Corp.*, 749 F. Supp. 1545, 1555 (D. Colo. 1990) (quoting *Brock* approvingly), *aff’d*, 972 F.2d 304 (10th Cir. 1992).

By contrast, a number of courts are more cautious about or reject using significance testing as a necessary condition, instead recognizing that assessing the likelihood of random error is important in determining the probative value of a study. In *Allen v. United States*, 588 F. Supp. 247, 417 (D. Utah 1984), the court stated, “The cold statement that a given relationship is not ‘statistically significant’ cannot be read to mean there is no probability of a relationship.” The Third Circuit described confidence intervals (i.e., the range of values that would be found in similar studies due to chance, with a specified level of confidence) and their use as an alternative to statistical significance in *DeLuca v. Merrell Dow Pharmaceuticals, Inc.*, 911 F.2d 941, 948–49 (3d Cir. 1990). See also *Milward v. Acuity Specialty Products Group, Inc.*, 639 F.3d 11, 24–25 (1st Cir. 2011) (recognizing the difficulty of obtaining statistically significant results when the disease under investigation occurs rarely and concluding that district court erred in imposing a statistical significance threshold); *Turpin v. Merrell Dow Pharms., Inc.*, 959 F.2d 1349, 1357 (6th Cir. 1992) (“The defendant’s claim overstates the persuasive power of these statistical studies. An analysis of this evidence demonstrates that it is possible that Bendectin causes birth defects even though these studies do not detect a significant association.”); *In re Viagra Prods. Liab. Litig.*, 572 F. Supp. 2d 1071, 1090 (D. Minn. 2008) (holding that, for purposes of supporting an opinion on general causation, a study does not have to find results with statistical significance); *United States v. Philip Morris USA, Inc.*, 449 F. Supp. 2d 1, 706 n.29 (D.D.C. 2006) (rejecting the position of an expert who denied that the causal connection between smoking and lung cancer had been established, in part, on the ground that any study that found an association that was not statistically significant must be excluded from consideration); *Cook v. Rockwell Int’l Corp.*, 580 F. Supp.

any study whose p -value is not less than the level chosen for statistical significance should be rejected as inadequate to disprove the null hypothesis. Others are critical of using strict significance testing, which rejects all studies with an observed p -value below that specified level. Epidemiologists have become increasingly sophisticated in addressing the issue of random error and examining the data from a study to ascertain what information they may provide about the relationship between an agent and a disease, without the necessity of rejecting all studies that are not statistically significant.⁸⁶ Meta-analysis, as well, a method for pooling the results of multiple studies, sometimes can ameliorate concerns about random error.⁸⁷

Calculation of a confidence interval permits a more refined assessment of appropriate inferences about the association found in an epidemiologic study.⁸⁸

2d 1071, 1103 (D. Colo. 2006) (“The statistical significance or insignificance of Dr. Clapp’s results may affect the weight given to his testimony, but does not determine its admissibility under Rule 702.”); *In re Ephedra Prods. Liab. Litig.*, 393 F. Supp. 2d 181, 186 (S.D.N.Y. 2005) (“[T]he absence of epidemiologic studies establishing an increased risk from ephedra of sufficient statistical significance to meet scientific standards of causality does not mean that the causality opinions of the PCC’s experts must be excluded entirely.”).

Although the trial court had relied in part on the absence of statistically significant epidemiologic studies, the Supreme Court in *Daubert v. Merrell Dow Pharmaceuticals, Inc.*, 509 U.S. 579 (1993), did not explicitly address the matter. The Court did, however, refer to “the known or potential rate of error” in identifying factors relevant to the scientific validity of an expert’s methodology. *Id.* at 594. The Court did not address any specific rate of error, although two cases that it cited affirmed the admissibility of voice spectrograph results that the courts reported were subject to a 2%–6% chance of error owing to either false matches or false eliminations. One commentator has concluded, “*Daubert* did not set a threshold level of statistical significance either for admissibility or for sufficiency of scientific evidence.” *Developments in the Law*, *supra* note 79, at 1535–36, 1540–46. The Supreme Court in *General Electric Co. v. Joiner*, 522 U.S. 136, 145–47 (1997), adverted to the lack of statistical significance in one study relied on by an expert as a ground for ruling that the district court had not abused its discretion in excluding the expert’s testimony.

In *Matrixx Initiatives, Inc. v. Siracusano*, 131 S. Ct. 1309 (2011), the Supreme Court was confronted with a question somewhat different from the relationship between statistically significant study results and causation. *Matrixx* was a securities fraud case in which the defendant argued that unless adverse event reports from use of a drug are statistically significant, the information about them is not material, as a matter of law (materiality is required as an element of a fraud claim). Defendant’s claim was premised on the idea that only statistically significant results can be a basis for an inference of causation. The Court, unanimously, rejected that claim, citing cases in which courts had permitted expert witnesses to testify to toxic causation in the absence of any statistically significant studies.

For a hypercritical assessment of statistical significance testing that nevertheless identifies much inappropriate overreliance on it, see Stephen T. Ziliak & Deidre N. McCloskey, *The Cult of Statistical Significance* (2008).

86. See Sanders, *supra* note 13, at 342 (describing the improved handling and reporting of statistical analysis in studies of Bendectin after 1980).

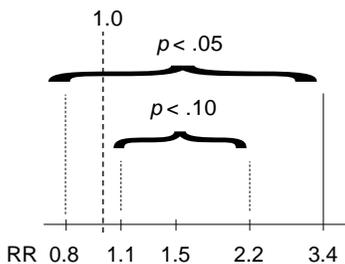
87. See *infra* Section VI.

88. Kenneth Rothman, Professor of Public Health at Boston University and Adjunct Professor of Epidemiology at the Harvard School of Public Health, is one of the leaders in advocating the use of confidence intervals and rejecting strict significance testing. In *DeLuca*, 911 F.2d at 947, the Third Circuit discussed Rothman’s views on the appropriate level of alpha and the use of con-

A confidence interval is a range of possible values calculated from the results of a study. If a 95% confidence interval is specified, the range encompasses the results we would expect 95% of the time if samples for new studies were repeatedly drawn from the same population. Thus, the width of the interval reflects random error.

The narrower the confidence interval, the more statistically stable the results of the study. The advantage of a confidence interval is that it displays more information than significance testing. “Statistically significant” does not convey the magnitude of the association found in the study or indicate how statistically stable that association is. A confidence interval shows the boundaries of the relative risk based on selected levels of alpha or statistical significance. Just as the p -value does not provide the probability that the risk estimate found in a study is correct, the confidence interval does not provide the range within which the true risk must lie. Rather, the confidence interval reveals the likely range of risk estimates consistent with random error. An example of two confidence intervals that might be calculated for a given relative risk is displayed in Figure 4.

Figure 4. Confidence intervals.



The confidence intervals shown in Figure 4 are for a study that found a relative risk of 1.5, with boundaries of 0.8 to 3.4 when the alpha is set to .05 (equivalently, a confidence level of .95), and with boundaries of 1.1 to 2.2 when alpha is set to .10 (equivalently, a confidence level of .90). The confidence interval for alpha equal to .10 is narrower because it encompasses only 90% of the expected test results.

By contrast, the confidence interval for alpha equal to .05 includes the expected outcomes for 95% of the tests. To generalize this point, the lower the alpha chosen (and therefore the more stringent the exclusion of possible random error) the wider the confidence interval. At a given alpha, the width of the confidence interval is

confidence intervals. In *Turpin*, 959 F.2d at 1353–54 n.1, the court discussed the relationship among confidence intervals, alpha, and power. See also *Cook v. Rockwell Int'l Corp.*, 580 F. Supp. 2d 1071, 1100–01 (D. Colo. 2006) (discussing confidence intervals, alpha, and significance testing). The use of confidence intervals in evaluating sampling error more generally than in the epidemiologic context is discussed in David H. Kaye & David A. Freedman, Reference Guide on Statistics, Section IV.A, in this manual.

determined by sample size. All other things being equal, the larger the sample size, the narrower the confidence boundaries (indicating greater numerical stability). For a given risk estimate, a narrower confidence interval reflects a decreased likelihood that the association found in the study would occur by chance if the true association is 1.0.⁸⁹

For the example in Figure 4, the boundaries of the confidence interval with alpha set at .05 encompass a relative risk of 1.0, and the result would be said to be not statistically significant at the .05 level. Alternatively, if the confidence boundaries are defined as an alpha equal to .10, then the confidence interval no longer includes a relative risk of 1.0, and the result would be characterized as statistically significant at the .10 level.

2. False negatives

As Figure 4 illustrates, false positives can be reduced by adopting more stringent values for alpha. Using an alpha of .05 will result in fewer false positives than using an alpha of .10, and an alpha of .01 or .001 would produce even fewer false positives.⁹⁰ The tradeoff for reducing false positives is an increase in false-negative errors (also called beta errors or Type II errors). This concept reflects the possibility that a study will be interpreted as “negative” (not disproving the null

89. Where multiple epidemiologic studies are available, a technique known as meta-analysis (see *infra* Section VI) may be used to combine the results of the studies to reduce the numerical instability of all the studies. See generally Diana B. Petitti, *Meta-analysis, Decision Analysis, and Cost-Effectiveness Analysis: Methods for Quantitative Synthesis in Medicine* (2d ed. 2000). Meta-analysis is better suited to combining results from randomly controlled experimental studies, but if carefully performed it may also be helpful for observational studies, such as those in the epidemiologic field. See Zachary B. Gerberg & Ralph I. Horwitz, *Resolving Conflicting Clinical Trials: Guidelines for Meta-Analysis*, 41 J. Clin. Epidemiol. 503 (1988). In *In re Bextra & Celebrex Marketing Sales Practices & Products Liability Litigation*, 524 F. Supp. 2d 1166 (N.D. Cal. 2007), the court relied on several meta-analyses of Celebrex at a 200-mg dose to conclude that the plaintiffs’ experts who proposed to testify to toxicity at that dosage failed to meet the requirements of *Daubert*. The court criticized those experts for the wholesale rejection of meta-analyses of observational studies.

In *In re Paoli Railroad Yard PCB Litigation*, 916 F.2d 829, 856–57 (3d Cir. 1990), the court discussed the use and admissibility of meta-analysis as a scientific technique. Overturning the district court’s exclusion of a report using meta-analysis, the Third Circuit observed that meta-analysis is a regularly used scientific technique. The court recognized that the technique might be poorly performed, and it required the district court to reconsider the validity of the expert’s work in performing the meta-analysis. See also *E.R. Squibb & Sons, Inc. v. Stuart Pharms.*, No. 90-1178, 1990 U.S. Dist. LEXIS 15788, at *41 (D.N.J. Oct. 16, 1990) (acknowledging the utility of meta-analysis but rejecting its use in that case because one of the two studies included was poorly performed); *Tobin v. Astra Pharm. Prods., Inc.*, 993 F.2d 528, 538–39 (6th Cir. 1992) (identifying an error in the performance of a meta-analysis, in which the Food and Drug Administration pooled data from control groups in different studies in which some gave the controls a placebo and others gave the controls an alternative treatment).

90. It is not uncommon in genome-wide association studies to set the alpha at .00001 or even lower because of the large number of associations tested in such studies. Reducing alpha is designed to limit the number of false-positive findings.

hypothesis), when in fact there is a true association of a specified magnitude.⁹¹ The beta for any study can be calculated only based on a specific alternative hypothesis about a given positive relative risk and a specific level of alpha selected.⁹²

3. Power

When a study fails to find a statistically significant association, an important question is whether the result tends to exonerate the agent's toxicity or is essentially inconclusive with regard to toxicity.⁹³ The concept of power can be helpful in evaluating whether a study's outcome is exonerative or inconclusive.⁹⁴

The power of a study is the probability of finding a statistically significant association of a given magnitude (if it exists) in light of the sample sizes used in the study. The power of a study depends on several factors: the sample size; the level of alpha (or statistical significance) specified; the background incidence of disease; and the specified relative risk that the researcher would like to detect.⁹⁵ Power curves can be constructed that show the likelihood of finding any given relative risk in light of these factors. Often, power curves are used in the design of a study to determine what size the study populations should be.⁹⁶

The power of a study is the complement of beta ($1 - \beta$). Thus, a study with a likelihood of .25 of failing to detect a true relative risk of 2.0⁹⁷ or greater has a power of .75. This means the study has a 75% chance of detecting a true relative risk of 2.0. If the power of a negative study to find a relative risk of 2.0 or greater

91. See also *DeLuca v. Merrell Dow Pharms., Inc.*, 911 F.2d 941, 947 (3d Cir. 1990).

92. See Green, *supra* note 47, at 684–89.

93. Even when a study or body of studies tends to exonerate an agent, that does not establish that the agent is absolutely safe. See *Cooley v. Lincoln Elec. Co.*, 693 F. Supp. 2d 767 (N.D. Ohio 2010). Epidemiology is not able to provide such evidence.

94. See Fienberg et al., *supra* note 72, at 22–23. Thus, in *Smith v. Wyeth-Ayerst Labs. Co.*, 278 F. Supp. 2d 684, 693 (W.D.N.C. 2003) and *Cooley v. Lincoln Electric Co.*, 693 F. Supp. 2d 767, 773 (N.D. Ohio 2010), the courts recognized that the power of a study was critical to assessing whether the failure of the study to find a statistically significant association was exonerative of the agent or inconclusive. See also *Procter & Gamble Pharms., Inc. v. Hoffmann-LaRoche Inc.*, No. 06 Civ. 0034(PAC), 2006 WL 2588002, at *32 n.16 (S.D.N.Y. Sept. 6, 2006) (discussing power curves and quoting the second edition of this reference guide); *In re Phenylpropanolamine (PPA) Prods. Liab. Litig.*, 289 F. Supp. 2d 1230, 1243–44 (W.D. Wash. 2003) (explaining expert's testimony that "statistical reassurance as to lack of an effect would require an upper bound of a reasonable confidence interval close to the null value"); *Ruff v. Ensign-Bickford Indus., Inc.*, 168 F. Supp. 2d 1271, 1281 (D. Utah 2001) (explaining why a study should be treated as inconclusive rather than exonerative based on small number of subjects in study).

95. See Malcolm Gladwell, *How Safe Are Your Breasts?* New Republic, Oct. 24, 1994, at 22, 26.

96. For examples of power curves, see Kenneth J. Rothman, *Modern Epidemiology* 80 (1986); Pagano & Gauvreau, *supra* note 59, at 245.

97. We use a relative risk of 2.0 for illustrative purposes because of the legal significance courts have attributed to this magnitude of association. See *infra* Section VII.

is low, it has substantially less probative value than a study with similar results but a higher power.⁹⁸

B. What Biases May Have Contributed to an Erroneous Association?

The second major reason for an invalid outcome in epidemiologic studies is systematic error or bias. Bias may arise in the design or conduct of a study, data collection, or data analysis. The meaning of scientific bias differs from conventional (and legal) usage, in which bias refers to a partisan point of view.⁹⁹ When scientists use the term *bias*, they refer to anything that results in a systematic (nonrandom) error in a study result and thereby compromises its validity. Two important categories of bias are selection bias (inappropriate methodology for selection of study subjects) and information bias (a flaw in measuring exposure or disease in the study groups).

Most epidemiologic studies have some degree of bias that may affect the outcome. If major bias is present, it may invalidate the study results. Finding the bias, however, can be difficult, if not impossible. In reviewing the validity of an epidemiologic study, the epidemiologist must identify potential biases and analyze the amount or kind of error that might have been induced by the bias. Often, the direction of error can be determined; depending on the specific type of bias, it may exaggerate the real association, dilute it, or even completely mask it.

1. Selection bias

Selection bias refers to the error in an observed association that results from the method of selection of cases and controls (in a case-control study) or exposed and unexposed individuals (in a cohort study).¹⁰⁰ The selection of an appropriate

98. See also David H. Kaye & David A. Freedman, Reference Guide on Statistics, Section IV.C.1, in this manual.

99. A Dictionary of Epidemiology 15 (John M. Last ed., 3d ed. 1995); Edmond A. Murphy, The Logic of Medicine 239–62 (1976).

100. Selection bias is defined as “[e]rror due to systematic differences in characteristics between those who are selected for study and those who are not.” A Dictionary of Epidemiology, *supra* note 98, at 153. In *In re “Agent Orange” Product Liability Litigation*, 597 F. Supp. 740, 783 (E.D.N.Y. 1985), *aff’d*, 818 F.2d 145 (2d Cir. 1987), the court expressed concern about selection bias. The exposed cohort consisted of young, healthy men who served in Vietnam. Comparing the mortality rate of the exposed cohort and that of a control group made up of civilians might have resulted in error that was a result of selection bias. Failing to account for health status as an independent variable tends to understate any association between exposure and disease in studies in which the exposed cohort is healthier. See also *In re Baycol Prods. Litig.*, 532 F. Supp. 2d 1029, 1043 (D. Minn. 2007) (upholding admissibility of testimony by expert witness who criticized study based on selection bias).

control group has been described as the Achilles' heel of a case-control study.¹⁰¹ Ideally, controls should be drawn from the same population that produced the cases. Selecting control participants becomes problematic if the control participants are selected for reasons that are related to their having the exposure being studied. For example, a study of the effect of smoking on heart disease will suffer selection bias if subjects of the study are volunteers and the decision to volunteer is affected by both being a smoker and having a family history of heart disease. The association will be biased upward because of the additional disease among the exposed smokers caused by genetics.

Hospital-based studies, which are relatively common among researchers located in medical centers, illustrate the problem. Suppose an association is found between coffee drinking and coronary heart disease in a study using hospital patients as controls. The problem is that the hospitalized control group may include individuals who had been advised against drinking coffee for medical reasons, such as to prevent the aggravation of a peptic ulcer. In other words, the controls may become eligible for the study because of their medical condition, which is in turn related to their exposure status—their likelihood of avoiding coffee. If this is true, the amount of coffee drinking in the control group would understate the extent of coffee drinking expected in people who do not have the disease, and thus bias upwardly (i.e., exaggerate) any odds ratio observed.¹⁰² Bias in hospital studies may also understate the true odds ratio when the exposures at issue led to the cases' hospitalizations and also contributed to the controls' chances of hospitalization.

Just as cases and controls in case-control studies should be selected independently of their exposure status, so the exposed and unexposed participants in cohort studies should be selected independently of their disease risk.¹⁰³ For example, if women with hysterectomies are overrepresented among exposed women in a cohort study of cervical cancer, this could overstate the association between the exposure and the disease.

A further source of selection bias occurs when those selected to participate decline to participate or drop out before the study is completed. Many studies have shown that individuals who participate in studies differ significantly from those who do not. If a significant portion of either study group declines to participate, the researcher should investigate whether those who declined are different from those who agreed. The researcher can compare relevant characteristics of those who

101. William B. Kannel & Thomas R. Dawber, *Coffee and Coronary Disease*, 289 *New Eng. J. Med.* 100 (1973) (editorial).

102. Hershel Jick et al., *Coffee and Myocardial Infarction*, 289 *New Eng. J. Med.* 63 (1973).

103. When unexposed controls may differ from the exposed cohort because exposure is associated with other risk (or protective factors), investigators can attempt to measure and adjust for those differences, as explained in Section IV.C.3, *infra*. See also Martha J. Radford & JoAnne M. Foody, *How Do Observational Studies Expand the Evidence Base for Therapy?* 286 *JAMA* 1228 (2001) (discussing the use of propensity analysis to adjust for potential confounding and selection biases that may occur from nonrandomization).

participate with those who do not to show the extent to which the two groups are comparable. Similarly, if a significant number of subjects drop out of a study before completion, the remaining subjects may not be representative of the original study populations. The researcher should examine whether that is the case.

The fact that a study may suffer from selection bias does not necessarily invalidate its results. A number of factors may suggest that a bias, if present, had only limited effect. If the association is particularly strong, for example, bias is less likely to account for all of it. In addition, a consistent association across different control groups suggests that possible biases applicable to a particular control group are not invalidating. Similarly, a dose–response relationship (*see* Section V.C, *infra*) found among multiple groups exposed to different doses of the agent would provide additional evidence that biases applicable to the exposed group are not a major problem.

2. Information bias

Information bias is a result of inaccurate information about either the disease or the exposure status of the study participants or a result of confounding. In a case-control study, potential information bias is an important consideration because the researcher depends on information from the past to determine exposure and disease and their temporal relationship.¹⁰⁴ In some situations, the researcher is required to interview the subjects about past exposures, thus relying on the subjects' memories. Research has shown that individuals with disease (cases) tend to recall past exposures more readily than individuals with no disease (controls);¹⁰⁵ this creates a potential for bias called recall bias.

For example, consider a case-control study conducted to examine the cause of congenital malformations. The epidemiologist is interested in whether the malformations were caused by an infection during the mother's pregnancy.¹⁰⁶ A group of mothers of malformed infants (cases) and a group of mothers of infants with no

104. Information bias can be a problem in cohort studies as well. When exposure is determined retrospectively, there can be a variety of impediments to obtaining accurate information. Similarly, when disease status is determined retrospectively, bias is a concern. The determination that asbestos is a cause of mesothelioma was hampered by inaccurate death certificates that identified lung cancer rather than mesothelioma, a rare form of cancer, as the cause of death. *See* I.J. Selikoff et al., *Mortality Experience of Insulation Workers in the United States and Canada*, 220 *Ann. N.Y. Acad. Sci.* 91, 110–11 (1979).

105. Steven S. Coughlin, *Recall Bias in Epidemiological Studies*, 43 *J. Clinical Epidemiology* 87 (1990).

106. *See* Brock v. Merrell Dow Pharms., Inc., 874 F.2d 307, 311–12 (5th Cir. 1989) (discussion of recall bias among women who bear children with birth defects). We note that the court was mistaken in its assertion that a confidence interval could correct for recall bias, or for any bias for that matter. Confidence intervals are a statistical device for analyzing error that may result from random sampling. Systematic errors (bias) in the design or data collection are not addressed by statistical methods, such as confidence intervals or statistical significance. *See* Green, *supra* note 47, at 667–68; Vincent M. Brannigan et al., *Risk, Statistical Inference, and the Law of Evidence: The Use of Epidemiological Data in Toxic Tort Cases*, 12 *Risk Analysis* 343, 344–45 (1992).

malformation (controls) are interviewed regarding infections during pregnancy. Mothers of children with malformations may recall an inconsequential fever or runny nose during pregnancy that readily would be forgotten by a mother who had a normal infant. Even if in reality the infection rate in mothers of malformed children is no different from the rate in mothers of normal children, the result in this study would be an apparently higher rate of infection in the mothers of the children with the malformations solely on the basis of recall differences between the two groups.¹⁰⁷ The issue of recall bias can sometimes be evaluated by finding an alternative source of data to validate the subject's response (e.g., blood test results from prenatal visits or medical records that document symptoms of infection).¹⁰⁸ Alternatively, the mothers' responses to questions about other exposures may shed light on the presence of a bias affecting the recall of the relevant exposures. Thus, if mothers of cases do not recall greater exposure than controls' mothers to pesticides, children with German measles, and so forth, then one can have greater confidence in their recall of illnesses.

Bias may also result from reliance on interviews with surrogates who are individuals other than the study subjects. This is often necessary when, for example, a subject (in a case-control study) has died of the disease under investigation or may be too ill to be interviewed.

There are many sources of information bias that affect the measure of exposure, including its intensity and duration. Exposure to the agent can be measured directly or indirectly.¹⁰⁹ Sometimes researchers use a biological marker as a direct measure of exposure to an agent—an alteration in tissue or body fluids that occurs as a result of an exposure and that can be detected in the laboratory. Biological markers, however, are only available for a small number of toxins and usually only reveal whether a person was exposed.¹¹⁰ Biological markers rarely help determine the intensity or duration of exposure.¹¹¹

107. Thus, in *Newman v. Motorola, Inc.*, 218 F. Supp. 2d 769, 778 (D. Md. 2002), the court considered a study of the effect of cell phone use on brain cancer and concluded that there was good reason to suspect that recall bias affected the results of the study, which found an association between cell phone use and cancers on the side of the head where the cell phone was used but no association between cell phone use and overall brain tumors.

108. Two researchers who used a case-control study to examine the association between congenital heart disease and the mother's use of drugs during pregnancy corroborated interview data with the mother's medical records. See Sally Zierler & Kenneth J. Rothman, *Congenital Heart Disease in Relation to Maternal Use of Bendectin and Other Drugs in Early Pregnancy*, 313 *New Eng. J. Med.* 347, 347-48 (1985).

109. See *In re Paoli R.R. Yard PCB Litig.*, No. 86-2229, 1992 U.S. Dist LEXIS 18430, at *9-*11 (E.D. Pa. Oct. 21, 1992) (discussing valid methods of determining exposure to chemicals).

110. See Gary E. Marchant, *Genetic Susceptibility and Biomarkers in Toxic Injury Litigation*, 41 *Jurimetrics J.* 67, 68, 73-74, 95-97 (2000) (explaining concept of biomarkers, how they might be used to provide evidence of exposure or dose, discussing cases in which biomarkers were invoked in an effort to prove exposure, and concluding, "biomarkers are likely to be increasingly relied on to demonstrate exposure").

111. There are different definitions of dose, but dose often refers to the intensity or magnitude of exposure multiplied by the time exposed. See *Sparks v. Owens-Illinois, Inc.*, 38 *Cal. Rptr.* 2d 739,

Monitoring devices also can be used to measure exposure directly but often are not available for exposures that have occurred in the past. For past exposures, epidemiologists often use indirect measures of exposure, such as interviewing workers and reviewing employment records. Thus, all those employed to install asbestos insulation may be treated as having been exposed to asbestos during the period that they were employed. However, there may be a wide variation of exposure within any job, and these measures may have limited applicability to a given individual.¹¹² If the agent of interest is a drug, medical or hospital records can be used to determine past exposure. Thus, retrospective studies, which are often used for occupational or environmental investigations, entail measurements of exposure that are usually less accurate than prospective studies or followup studies, including ones in which a drug or medical intervention is the independent variable being measured.

742 (Ct. App. 1995). Other definitions of dose may be more appropriate in light of the biological mechanism of the disease.

For a discussion of the difficulties of determining dose from atomic fallout, see *Allen v. United States*, 588 F. Supp. 247, 425–26 (D. Utah 1984), *rev'd on other grounds*, 816 F.2d 1417 (10th Cir. 1987). The timing of exposure may also be critical, especially if the disease of interest is a birth defect. In *Smith v. Ortho Pharmaceutical Corp.*, 770 F. Supp. 1561, 1577 (N.D. Ga. 1991), the court criticized a study for its inadequate measure of exposure to spermicides. The researchers had defined exposure as receipt of a prescription for spermicide within 600 days of delivery, but this definition of exposure is too broad because environmental agents are likely to cause birth defects only during a narrow band of time.

A different, but related, problem often arises in court. Determining the plaintiff's exposure to the alleged toxic substance always involves a retrospective determination and may involve difficulties similar to those faced by an epidemiologist planning a study. Thus, in *John's Heating Service v. Lamb*, 46 P.3d 1024 (Alaska 2002), plaintiffs were exposed to carbon monoxide because of defendants' negligence with respect to a home furnace. The court observed: "[W]hile precise information concerning the exposure necessary to cause specific harm to humans and exact details pertaining to the plaintiff's exposure are beneficial, such evidence is not always available, or necessary, to demonstrate that a substance is toxic to humans given substantial exposure and need not invariably provide the basis for an expert's opinion on causation." *Id.* at 1035 (quoting *Westberry v. Gislaved Gummi AB*, 178 F.3d 257, 264 (4th Cir. 1999)); see also *Alder v. Bayer Corp.*, AGFA Div., 61 P.3d 1068, 1086–88 (Utah 2002) (summarizing other decisions on the precision with which plaintiffs must establish the dosage to which they were exposed). See generally Restatement (Third) of Torts: Liability for Physical and Emotional Harm § 28 cmt. c(2) & rptrs. note (2010).

In asbestos litigation, a number of courts have adopted a requirement that the plaintiff demonstrate (1) regular use by an employer of the defendant's asbestos-containing product, (2) the plaintiff's proximity to that product, and (3) exposure over an extended period of time. See, e.g., *Lohrmann v. Pittsburgh Corning Corp.*, 782 F.2d 1156, 1162–64 (4th Cir. 1986); *Gregg v. V-J Auto Parts, Inc.*, 943 A.2d 216, 226 (Pa. 2007).

112. Frequently, occupational epidemiologists employ study designs that consider all agents to which those who work in a particular occupation are exposed because they are trying to determine the hazards associated with that occupation. Isolating one of the agents for examination would be difficult if not impossible. These studies, then, present difficulties when employed in court in support of a claim by a plaintiff who was exposed to only one or fewer than all of the agents present at the worksite that was the subject of the study. See, e.g., *Knight v. Kirby Inland Marine Inc.*, 482 F.3d 347, 352–53 (5th Cir. 2007) (concluding that case-control studies of cancer that entailed exposure to a variety of organic solvents at job sites did not support claims of plaintiffs who claimed exposure to benzene caused their cancers).

The route (e.g., inhalation or absorption), duration, and intensity of exposure are important factors in assessing disease causation. Even with environmental monitoring, the dose measured in the environment generally is not the same as the dose that reaches internal target organs. If the researcher has calculated the internal dose of exposure, the scientific basis for this calculation should be examined for soundness.¹¹³

In assessing whether the data may reflect inaccurate information, one must assess whether the data were collected from objective and reliable sources. Medical records, government documents, employment records, death certificates, and interviews are examples of data sources that are used by epidemiologists to measure both exposure and disease status.¹¹⁴ The accuracy of a particular source may affect the validity of a research finding. If different data sources are used to collect information about a study group, differences in the accuracy of those sources may affect the validity of the findings. For example, using employment records to gather information about exposure to narcotics probably would lead to inaccurate results, because employees tend to keep such information private. If the researcher uses an unreliable source of data, the study may not be useful.

The kinds of quality control procedures used may affect the accuracy of the data. For data collected by interview, quality control procedures should probe the reliability of the individual and whether the information is verified by other sources. For data collected and analyzed in the laboratory, quality control procedures should probe the validity and reliability of the laboratory test.

Information bias may also result from inaccurate measurement of disease status. The quality and sophistication of the diagnostic methods used to detect a disease should be assessed.¹¹⁵ The proportion of subjects who were examined also should be questioned. If, for example, many of the subjects refused to be tested, the fact that the test used was of high quality would be of relatively little value.

113. See also Bernard D. Goldstein & Mary Sue Henifin, Reference Guide on Toxicology, Section I.D, in this manual.

114. Even these sources may produce unanticipated error. Identifying the causal connection between asbestos and mesothelioma, a rare form of cancer, was complicated and delayed because doctors who were unfamiliar with mesothelioma erroneously identified other causes of death in death certificates. See David E. Lilienfeld & Paul D. Gunderson, *The "Missing Cases" of Pleural Malignant Mesothelioma in Minnesota, 1979–81: Preliminary Report*, 101 Pub. Health Rep. 395, 397–98 (1986).

115. The hazards of adversarial review of epidemiologic studies to determine bias is highlighted by *O'Neill v. Novartis Consumer Health, Inc.*, 55 Cal. Rptr. 3d 551, 558–60 (Ct. App. 2007). Defendant's experts criticized a case-control study relied on by plaintiff on the ground that there was misclassification of exposure status among the cases. Plaintiff objected to this criticism because defendant's experts had only examined the cases for exposure misclassification, which would tend to exaggerate any association by providing an inaccurately inflated measure of exposure in the cases. The experts failed to examine whether there was misclassification in the controls, which, if it existed, would tend to incorrectly diminish any association.

The scientific validity of the research findings is influenced by the reliability of the diagnosis of disease or health status under study.¹¹⁶ The disease must be one that is recognized and defined to enable accurate diagnoses.¹¹⁷ Subjects' health status may be essential to the hypothesis under investigation. For example, a researcher interested in studying spontaneous abortion in the first trimester must determine that study subjects are pregnant. Diagnostic criteria that are accepted by the medical community should be used to make the diagnosis. If a diagnosis had been made at a time when home pregnancy kits were known to have a high rate of false-positive results (indicating pregnancy when the woman is not pregnant), the study will overestimate the number of spontaneous abortions.

Misclassification bias is a consequence of information bias in which, because of problems with the information available, individuals in the study may be misclassified with regard to exposure status or disease status. Bias due to exposure misclassification can be differential or nondifferential. In nondifferential misclassification, the inaccuracies in determining exposure are independent of disease status, or the inaccuracies in diagnoses are independent of exposure status—in other words, the data are crude, with a great deal of random error. This is a common problem. Generally, nondifferential misclassification bias leads to a shift in the odds ratio toward one, or, in other words, toward a finding of no effect. Thus, if the errors are nondifferential, it is generally misguided to criticize an apparent association between an exposure and disease on the ground that data were inaccurately classified. Instead, nondifferential misclassification generally underestimates the true size of the association.

Differential misclassification is systematic error in determining exposure in cases as compared with controls, or disease status in unexposed cohorts relative to exposed cohorts. In a case-control study this would occur, for example, if, in the

116. In *In re Swine Flu Immunization Products Liability Litigation*, 508 F. Supp. 897, 903 (D. Colo. 1981), *aff'd sub nom. Lima v. United States*, 708 F.2d 502 (10th Cir. 1983), the court critically evaluated a study relied on by an expert whose testimony was stricken. In that study, determination of whether a patient had Guillain-Barré syndrome was made by medical clerks, not physicians who were familiar with diagnostic criteria.

117. The difficulty of ill-defined diseases arose in some of the silicone gel breast implant cases. Thus, in *Grant v. Bristol-Myers Squibb*, 97 F. Supp. 2d 986 (D. Ariz. 2000), in the face of a substantial body of exonerative epidemiologic evidence, the female plaintiff alleged she suffered from an atypical systemic joint disease. The court concluded:

As a whole, the Court finds that the evidence regarding systemic disease as proposed by Plaintiffs' experts is not scientifically valid and therefore will not assist the trier of fact. As for the atypical syndrome that is suggested, where experts propose that breast implants cause a disease but cannot specify the criteria for diagnosing the disease, it is incapable of epidemiologic testing. This renders the experts' methods insufficiently reliable to help the jury.

Id. at 992; *see also* *Burton v. Wyeth-Ayerst Labs.*, 513 F. Supp. 2d 719, 722–24 (N.D. Tex. 2007) (parties disputed whether cardiology problem involved two separate diseases or only one; court concluded that all experts in the case reflected a view that there was but a single disease); *In re Breast Implant Cases*, 942 F. Supp. 958, 961 (E.D.N.Y. & S.D.N.Y. 1996).

process of anguishing over the possible causes of the disease, parents of ill children recalled more exposures to a particular agent than actually occurred, or if parents of the controls, for whom the issue was less emotionally charged, recalled fewer. This can also occur in a cohort study in which, for example, birth control users (the exposed cohort) are monitored more closely for potential side effects, leading to a higher rate of disease identification in that cohort than in the unexposed cohort. Depending on how the misclassification occurs, a differential bias can produce an error in either direction—the exaggeration or understatement of a true association.

3. Other conceptual problems

There are dozens of other potential biases that can occur in observational studies, which is an important reason why clinical studies (when ethical) are often preferable. Sometimes studies are limited by flawed definitions or premises. For example, if the researcher defines the disease of interest as all birth defects, rather than a specific birth defect, there should be a scientific basis to hypothesize that the effects of the agent being investigated could be so broad. If the effect is in fact more limited, the result of this conceptualization error could be to dilute or mask any real effect that the agent might have on a specific type of birth defect.¹¹⁸

Some biases go beyond errors in individual studies and affect the overall body of available evidence in a way that skews what appears to be the universe of evidence. Publication bias is the tendency for medical journals to prefer studies that find an effect.¹¹⁹ If negative studies are never published, the published literature will be biased. Financial conflicts of interest by researchers and the source of funding of studies have been shown to have an effect on the outcomes of such studies.¹²⁰

118. In *Brock v. Merrell Dow Pharmaceuticals, Inc.*, 874 F.2d 307, 312 (5th Cir. 1989), the court discussed a reanalysis of a study in which the effect was narrowed from all congenital malformations to limb reduction defects. The magnitude of the association changed by 50% when the effect was defined in this narrower fashion. See Rothman et al. *supra* note 61, at 144 (“Unwarranted assurances of a lack of any effect can easily emerge from studies in which a wide range of etiologically unrelated outcomes are grouped.”).

119. Investigators may contribute to this effect by neglecting to submit negative studies for publication.

120. See Jerome P. Kassirer, On the Take: How Medicine’s Complicity with Big Business Can Endanger Your Health 79–84 (2005); J.E. Bekelman et al., *Scope and Impact of Financial Conflicts of Interest in Biomedical Research: A Systematic Review*, 289 JAMA 454 (2003). Richard Smith, the editor in chief of the British Medical Journal, wrote on this subject:

The major determinant of whether reviews of passive smoking concluded it was harmful was whether the authors had financial ties with tobacco manufacturers. In the disputed topic of whether third-generation contraceptive pills cause an increase in thromboembolic disease, studies funded by the pharmaceutical industry find that they don’t and studies funded by public money find that they do.

Richard Smith, *Making Progress with Competing Interests*, 325 Brit. Med. J. 1375, 1376 (2002).

Examining a study for potential sources of bias is an important task that helps determine the accuracy of a study's conclusions. In addition, when a source of bias is identified, it may be possible to determine whether the error tended to exaggerate or understate the true association. Thus, bias may exist in a study that nevertheless has probative value.

Even if one concludes that the findings of a study are statistically stable and that biases have not created significant error, additional considerations remain. As repeatedly noted, an association does not necessarily mean a causal relationship exists. To make a judgment about causation, a knowledgeable expert¹²¹ must consider the possibility of confounding factors. The expert must also evaluate several criteria to determine whether an inference of causation is appropriate.¹²² These matters are discussed below.

C. *Could a Confounding Factor Be Responsible for the Study Result?*¹²³

The third major reason for error in epidemiologic studies is confounding. Confounding occurs when another causal factor (the confounder) confuses the relationship between the agent of interest and outcome of interest.¹²⁴ (Confounding and selection bias (Section IV.B.1, *supra*) can, depending on terminology, overlap.) Thus, one instance of confounding is when a confounder is both a risk factor for the disease and a factor associated with the exposure of interest. For example, researchers may conduct a study that finds individuals with gray hair have a higher rate of death than those with hair of another color. Instead of hair color having an impact on death, the results might be explained by the confounding factor of age. If old age is associated differentially with the gray-haired group (those with gray hair tend to be older), old age may be responsible for the association found between hair color and death.¹²⁵ Researchers must separate the relationship between gray hair and risk of death from that of old age and risk of death. When researchers find an association between an agent and a disease, it is critical to determine whether the association is causal or the result of confounding.¹²⁶ Some

121. In a lawsuit, this would be done by an expert. In science, the effort is usually conducted by a panel of experts.

122. For an excellent example of the authors of a study analyzing whether an inference of causation is appropriate in a case-control study examining whether bromocriptine (Parlodel)—a lactation suppressant—causes seizures in postpartum women, see Kenneth J. Rothman et al., *Bromocriptine and Puerpal Seizures*, 1 *Epidemiology* 232, 236–38 (1990).

123. See *Grassis v. Johns-Manville Corp.*, 591 A.2d 671, 675 (N.J. Super. Ct. App. Div. 1991) (discussing the possibility that confounders may lead to an erroneous inference of a causal relationship).

124. See Rothman et al., *supra* note 61, at 129.

125. This example is drawn from Kahn & Sempos, *supra* note 31, at 63.

126. Confounding can bias a study result by either exaggerating or diluting any true association. One example of a confounding factor that may result in a study's outcome understating an

epidemiologists classify confounding as a form of bias. However, confounding is a reality—that is, the observed association of a factor and a disease is actually the result of an association with a third, confounding factor.¹²⁷

Confounding can be illustrated by a hypothetical prospective cohort study of the role of alcohol consumption and emphysema. The study is designed to investigate whether drinking alcohol is associated with emphysema. Participants are followed for a period of 20 years and the incidence of emphysema in the “exposed” (participants who consume more than 15 drinks per week) and the unexposed is compared. At the conclusion of the study, the relative risk of emphysema in the drinking group is found to be 2.0, an association that suggests a possible effect). But does this association reflect a true causal relationship or might it be the product of confounding?

One possibility for a confounding factor is smoking, a known causal risk factor for emphysema. If those who drink alcohol are more likely to be smokers than those who do not drink, then smoking may be responsible for some or all of the higher level of emphysema among those who do not drink.

A serious problem in observational studies such as this hypothetical study is that the individuals are not assigned randomly to the groups being compared.¹²⁸ As discussed above, randomization maximizes the possibility that exposures other than the one under study are evenly distributed between the exposed and the control cohorts.¹²⁹ In observational studies, by contrast, other forces, including self-selection, determine who is exposed to other (possibly causal) factors. The lack of randomization leads to the potential problem of confounding. Thus, for example, the exposed cohort might consist of those who are exposed at work to an agent suspected of being an industrial toxin. The members of this cohort may, however, differ from unexposed controls by residence, socioeconomic or health status, age, or other extraneous factors.¹³⁰ These other factors may be causing (or

association is vaccination. Thus, if a group exposed to an agent has a higher rate of vaccination for the disease under study than the unexposed group, the vaccination may reduce the rate of disease in the exposed group, thereby producing an association that is less than the true association without the confounding of vaccination.

127. *Schwab v. Philip Morris USA, Inc.*, 449 F. Supp. 2d 992, 1199–1200 (E.D.N.Y. 2006), *rev'd on other grounds*, 522 F.3d 215 (2d Cir. 2008), describes confounding that led to premature conclusions that low-tar cigarettes were safer than regular cigarettes. Smokers who chose to switch to low-tar cigarettes were different from other smokers in that they were more health conscious in other aspects of their lifestyles. Failure to account for that confounding—and measuring a healthy lifestyle is difficult even if it is identified as a potential confounder—biased the results of those studies.

128. Randomization attempts to ensure that the presence of a characteristic, such as coffee drinking, is governed by chance, as opposed to being determined by the presence of an underlying medical condition.

129. See Rothman et al., *supra* note 61, at 129; see also *supra* Section II.A.

130. See, e.g., *In re “Agent Orange” Prod. Liab. Litig.*, 597 F. Supp. 740, 783 (E.D.N.Y. 1984) (discussing the problem of confounding that might result in a study of the effect of exposure to Agent Orange on Vietnam servicemen), *aff'd*, 818 F.2d 145 (2d Cir. 1987).

protecting against) the disease, but because of potential confounding, an apparent (yet false) association of the disease with exposure to the agent may appear. Confounders, like smoking in the alcohol drinking study, do not reflect an error made by the investigators; rather, they reflect the inherently “uncontrolled” nature of exposure designations in observational studies. When they can be identified, confounders should be taken into account. Unanticipated confounding factors that are suspected after data collection can sometimes be controlled during data analysis, if data have been gathered about them.

To evaluate whether smoking is a confounding factor, the researcher would stratify each of the exposed and control groups into smoking and nonsmoking subgroups to examine whether subjects’ smoking status affects the study results. If the relationship between alcohol drinking and emphysema in the smoking subgroups is the same as that in the all-subjects group, smoking is not a confounding factor. If the subjects’ smoking status affects the relationship between drinking and emphysema, then smoking is a confounder, for which adjustment is required. If the association between drinking and emphysema completely disappears when the subjects’ smoking status is considered, then smoking is a confounder that fully accounts for the association with drinking observed. Table 4 reveals our hypothetical study’s results, with smoking being a confounding factor, which, when accounted for, eliminates the association. Thus, in the full cohort, drinkers have twice the risk of emphysema compared with nondrinkers. When the relationship between drinking and emphysema is examined separately in smokers and in nonsmokers, the risk of emphysema in drinkers compared with nondrinkers is not elevated in smokers or in nonsmokers. This is because smokers are disproportionately drinkers and have a higher rate of emphysema than nonsmokers. Thus, the relationship between drinking and emphysema in the full cohort is distorted by failing to take into account the relationship between being a drinker and a smoker.

Even after accounting for the effect of smoking, there is always a risk that an undiscovered or unrecognized confounding factor may contribute to a study’s findings, by either magnifying or reducing the observed association.¹³¹ It is, however, necessary to keep that risk in perspective. Often the mere possibility of uncontrolled confounding is used to call into question the results of a study. This was certainly the strategy of some seeking, or unwittingly helping, to undermine the implications of the studies persuasively linking cigarette smoking to lung cancer. The critical question is whether it is plausible that the findings of a given study could indeed be due to unrecognized confounders.

In designing a study, researchers sometimes make assumptions that cannot be validated or evaluated empirically. Thus, researchers may assume that a missing potential confounder is not needed for the analysis or that a variable used was adequately classified. Researchers employ a sensitivity analysis to assess the effect of those assumptions should they be incorrect. Conducting a sensitivity analysis

131. Rothman et al., *supra* note 61, at 129; *see also supra* Section II.A.

Table 4. Hypothetical Emphysema Study Data^a

Drinking Status	Total Cohort				Smokers				Nonsmokers			
	Total	Cases	Incidence	RR	Total	Cases	Incidence	RR	Total	Cases	Incidence	RR
Nondrinkers	471	16	0.034	1.0 ^b	111	9	0.081	1.0 ^b	360	7	0.019	1.0 ^b
Drinkers	739	41	0.069	2.0	592	48	0.081	1.0	147	3	0.020	1.0

^a The incidence of disease is not normally presented in an epidemiologic study, but we include it here to aid in comprehension of the ideas discussed in the text. ^b RR = relative risk. The relative risk for each of the cohorts is determined based on reference to the risk among nondrinkers; that is, the incidence of disease among drinkers is compared with nondrinkers for each of the three cohorts separately.

entails repeating the analysis using different assumptions (e.g., alternative corrections for missing data or for classifying data) to see if the results are sensitive to the varying assumptions. Such analyses can show that the assumptions are not likely to affect the findings or that alternative explanations cannot be ruled out.¹³²

1. What techniques can be used to prevent or limit confounding?

Choices in the design of a research project (e.g., methods for selecting the subjects) can prevent or limit confounding. In designing a study, the researcher must determine other risk factors for the disease under study. When a factor or factors, such as age, sex, or even smoking status, are risk factors and potential confounders in a study, investigators can limit the differential distribution of these factors in the study groups by selecting controls to “match” cases (or the exposed group) in terms of these variables. If the two groups are matched, for example, by age, then any association observed in the study cannot be due to age, the matched variable.¹³³

Restricting the persons who are permitted as subjects in a study is another method to control for confounders. If age or sex is suspected as a confounder, then the subjects enrolled in a study can be limited to those of one sex and those who are within a specified age range. When there is no variance among subjects in a study with regard to a potential confounder, confounding as a result of that variable is eliminated.

2. What techniques can be used to identify confounding factors?

Once the study data are ready to be analyzed, the researcher must assess a range of factors that could influence risk. In the hypothetical study, the researcher would evaluate whether smoking is a confounding factor by comparing the incidence of emphysema in smoking alcohol drinkers with the incidence in nonsmoking alcohol drinkers. If the incidence is substantially the same, smoking is not a confounding factor (e.g., smoking does not distort the relationship between alcohol drinking and the development of emphysema). If the incidence is substantially different, but still exists in the nonsmoking group, then smoking is a confounder, but does not wholly account for the association with alcohol drinking. If the association disappears, then smoking is a confounder that fully accounts for the association observed.

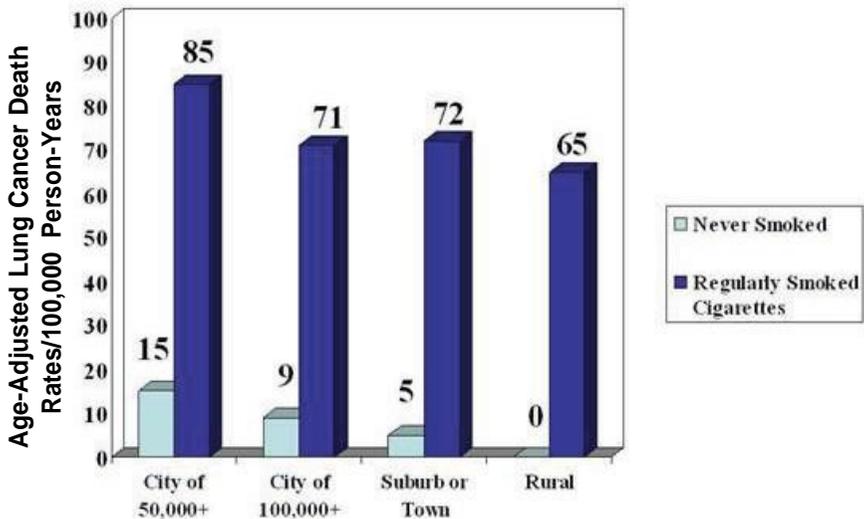
132. Kenneth Rothman & Sander Greenland, *Modern Epidemiology* (2d ed. 1998).

133. Selecting a control population based on matched variables necessarily affects the representativeness of the selected controls and may affect how generalizable the study results are to the population at large. However, for a study to have merit, it must first be internally valid; that is, it must not be subject to unreasonable sources of bias or confounding. Only after a study has been shown to meet this standard does its universal applicability or generalizability to the population at large become an issue. When a study population is not representative of the general or target population, existing scientific knowledge may permit reasonable inferences about the study's broader applicability, or additional confirmatory studies of other populations may be necessary.

3. What techniques can be used to control for confounding factors?

A good study design will consider potential confounders and obtain data about them if possible. If researchers have good data on potential confounders, they can control for those confounders in the data analysis. There are several analytic approaches to account for the distorting effects of a confounder, including stratification or multivariate analysis. Stratification permits an investigator to evaluate the effect of a suspected confounder by subdividing the study groups based on a confounding factor. Thus, in Table 4, drinkers have been stratified based on whether they smoke (the suspected confounder). To take another example that entails a continuous rather than dichotomous potential confounder, let us say we are interested in the relationship between smoking and lung cancer but suspect that air pollution or urbanization may confound the relationship. Thus, an observed relationship between smoking and lung cancer could theoretically be due in part to pollution, if smoking were more common in polluted areas. We could address this issue by stratifying our data by degree of urbanization and look at the relationship between smoking and lung cancer in each urbanization stratum. Figure 5 shows actual age-adjusted lung cancer mortality rates per 100,000 person-years by urban or rural classification and smoking category.¹³⁴

Figure 5: Age-adjusted lung cancer mortality rates per 100,000 person-years by urban or rural classification and smoking category.



Source: Adapted from E. Cuyler Hammond & Daniel Horn, *Smoking and Death Rates—Report on Forty-Four Months of Follow-Up of 187,783 Men: II, Death Rates by Cause*, 166 JAMA 1294 (1958).

134. This example and Figure 4 are from Leon Gordis, *Epidemiology* 254 (4th ed. 2009).

For each degree of urbanization, lung cancer mortality rates in smokers are shown by the dark gray bars, and nonsmoker mortality rates are indicated by light gray bars. From these data we see that in every level (or stratum) of urbanization, lung cancer mortality is higher in smokers than in nonsmokers. Therefore, the observed association of smoking and lung cancer cannot be attributed to level of urbanization. By examining each stratum separately, we, in effect, hold urbanization constant, and still find much higher lung cancer mortality in smokers than in nonsmokers.

For each degree of urbanization, lung cancer mortality rates and smokers are shown by the dark-colored bars, and nonsmoker mortality rates are indicated by light-colored bars. For these data we see that in every level (or stratum) of urbanization, lung cancer mortality is higher in smokers than in nonsmokers. Therefore, the observed association of lung cancer cannot be attributed to level of urbanization. By examining each stratum separately, we are, in effect, holding urbanization constant, and we still find much higher lung cancer mortality in smokers than in nonsmokers.

Multivariate analysis controls for the confounding factor through mathematical modeling. Models are developed to describe the simultaneous effect of exposure and confounding factors on the increase in risk.¹³⁵

Both of these methods allow for adjustment of the effect of confounders. They both modify an observed association to take into account the effect of risk factors that are not the subject of the study and that may distort the association between the exposure being studied and the disease outcomes. If the association between exposure and disease remains after the researcher completes the assessment and adjustment for confounding factors, the researcher must then assess whether an inference of causation is justified. This entails consideration of the Hill factors explained in Section V, *infra*.

V. General Causation: Is an Exposure a Cause of the Disease?

Once an association has been found between exposure to an agent and development of a disease, researchers consider whether the association reflects a true cause–effect relationship. When epidemiologists evaluate whether a cause–effect relationship exists between an agent and disease, they are using the term causation in a way similar to, but not identical to, the way that the familiar “but for,” or *sine qua non*, test is used in law for cause in fact. “Conduct is a factual cause of

135. For a more complete discussion of multivariate analysis, see Daniel L. Rubinfield, Reference Guide on Multiple Regression, in this manual.

[harm] when the harm would not have occurred absent the conduct.”¹³⁶ This is equivalent to describing the conduct as a necessary link in a chain of events that results in the particular event.¹³⁷ Epidemiologists use causation to mean that an increase in the incidence of disease among the exposed subjects would not have occurred had they not been exposed to the agent.¹³⁸ Thus, exposure is a necessary condition for the increase in the incidence of disease among those exposed.¹³⁹ The relationship between the epidemiologic concept of cause and the legal question of whether exposure to an agent caused an individual’s disease is addressed in Section VII.

As mentioned in Section I, epidemiology cannot prove causation; rather, causation is a judgment for epidemiologists and others interpreting the epidemiologic data.¹⁴⁰ Moreover, scientific determinations of causation are inherently tentative. The scientific enterprise must always remain open to reassessing the validity of past judgments as new evidence develops.

In assessing causation, researchers first look for alternative explanations for the association, such as bias or confounding factors, which are discussed in Section IV, *supra*. Once this process is completed, researchers consider how guidelines for inferring causation from an association apply to the available evidence. We emphasize that these guidelines are employed only *after* a study finds an association

136. Restatement (Third) of Torts: Liability for Physical and Emotional Harm § 26 (2010); *see also* Dan B. Dobbs, *The Law of Torts* § 168, at 409–11 (2000). When multiple causes are each operating and capable of causing an event, the but-for, or necessary-condition, concept for causation is problematic. This is the familiar “two-fires” scenario in which two independent fires simultaneously burn down a house and is sometimes referred to as overdetermined outcomes. Neither fire is a but-for, or necessary condition, for the destruction of the house, because either fire would have destroyed the house. *See* Restatement (Third) of Torts: Liability for Physical and Emotional Harm § 28 (2010). This two-fires situation is analogous to an individual being exposed to two agents, each of which is capable of causing the disease contracted by the individual. *See* *Basko v. Sterling Drug, Inc.*, 416 F.2d 417 (2d Cir. 1969). A difference between the disease scenario and the fire scenario is that, in the former, one will have no more than a probabilistic assessment of whether each of the exposures would have caused the disease in the individual.

137. *See supra* note 7; *see also* Restatement (Third) of Torts: Liability for Physical and Emotional Harm § 26 cmt. c (2010) (employing a “causal set” model to explain multiple elements, each of which is required for an outcome).

138. “The imputed causal association is at the group level, and does not indicate the cause of disease in individual subjects.” Bruce G. Charlton, *Attribution of Causation in Epidemiology: Chain or Mosaic?* 49 *J. Clin. Epidemiology* 105, 105 (1999).

139. *See* Rothman et al., *supra* note 61, at 8 (“We can define a cause of a specific disease event as an antecedent event, condition, or characteristic that was necessary for the occurrence of the disease at the moment it occurred, given that other conditions are fixed.”); *Allen v. United States*, 588 F. Supp. 247, 405 (D. Utah 1984) (quoting a physician on the meaning of the statement that radiation causes cancer), *rev’d on other grounds*, 816 F.2d 1417 (10th Cir. 1987).

140. Restatement (Third) of Torts: Liability for Physical and Emotional Harm § 28 cmt. c (2010) (“[A]n evaluation of data and scientific evidence to determine whether an inference of causation is appropriate requires judgment and interpretation.”).

to determine whether that association reflects a true causal relationship.¹⁴¹ These guidelines consist of several key inquiries that assist researchers in making a judgment about causation.¹⁴² Generally, researchers are conservative when it comes to assessing causal relationships, often calling for stronger evidence and more research before a conclusion of causation is drawn.¹⁴³

The factors that guide epidemiologists in making judgments about causation (and there is no threshold number that must exist) are¹⁴⁴

141. In a number of cases, experts attempted to use these guidelines to support the existence of causation in the absence of any epidemiologic studies finding an association. *See, e.g.,* *Rains v. PPG Indus., Inc.*, 361 F. Supp. 2d 829, 836–37 (S.D. Ill. 2004) (explaining Hill criteria and proceeding to apply them even though there was no epidemiologic study that found an association); *Soldo v. Sandoz Pharms. Corp.*, 244 F. Supp. 2d 434, 460–61 (W.D. Pa. 2003). There may be some logic to that effort, but it does not reflect accepted epidemiologic methodology. *See In re Fosamax Prods. Liab. Litig.*, 645 F. Supp. 2d 164, 187–88 (S.D.N.Y. 2009); *Dunn v. Sandoz Pharms. Corp.*, 275 F. Supp. 2d 672, 678–79 (M.D.N.C. 2003) (“The greater weight of authority supports Sandoz’ assertion that [use of] the Bradford Hill criteria is a method for determining whether the results of an epidemiologic study can be said to demonstrate causation and not a method for testing an unproven hypothesis.”); *Soldo*, 244 F. Supp. 2d at 514 (the Hill criteria “were developed as a mean[s] of interpreting an established association based on a body of epidemiologic research for the purpose of trying to judge whether the observed association reflects a causal relation between an exposure and disease.” (quoting report of court-appointed expert)).

142. *See* Mervyn Susser, *Causal Thinking in the Health Sciences: Concepts and Strategies in Epidemiology* (1973); *Gannon v. United States*, 571 F. Supp. 2d 615, 624 (E.D. Pa. 2007) (quoting expert who testified that the Hill criteria are “well-recognized” and widely used in the science community to assess general causation”); *Chapin v. A & L Parts, Inc.*, 732 N.W.2d 578, 584 (Mich. Ct. App. 2007) (expert testified that Hill criteria are the most well-utilized method for determining if an association is causal).

143. *Berry v. CSX Transp., Inc.*, 709 So. 2d 552, 568 n.12 (Fla. Dist. Ct. App. 1998) (“Almost all genres of research articles in the medical and behavioral sciences conclude their discussion with qualifying statements such as ‘there is still much to be learned.’ This is not, as might be assumed, an expression of ignorance, but rather an expression that all scientific fields are open-ended and can progress from their present state. . . .”); *Hall v. Baxter Healthcare Corp.*, 947 F. Supp. 1387 app. B. at 1446–51 (D. Or. 1996) (report of Merwyn R. Greenlick, court-appointed epidemiologist). In *Cadarian v. Merrell Dow Pharmaceuticals, Inc.*, 745 F. Supp. 409 (E.D. Mich. 1989), the court refused to permit an expert to rely on a study that the authors had concluded should not be used to support an inference of causation in the absence of independent confirmatory studies. The court did not address the question whether the degree of certainty used by epidemiologists before making a conclusion of cause was consistent with the legal standard. *See DeLuca v. Merrell Dow Pharms., Inc.*, 911 F.2d 941, 957 (3d Cir. 1990) (standard of proof for scientific community is not necessarily appropriate standard for expert opinion in civil litigation); *Wells v. Ortho Pharm. Corp.*, 788 F.2d 741, 745 (11th Cir. 1986).

144. *See Cook v. Rockwell Int’l Corp.*, 580 F. Supp. 2d 1071, 1098 (D. Colo. 2006) (“Defendants cite no authority, scientific or legal, that compliance with all, or even one, of these factors is required. . . . The scientific consensus is, in fact, to the contrary. It identifies Defendants’ list of factors as some of the nine factors or lenses that guide epidemiologists in making judgments about causation. . . . These factors are not tests for determining the reliability of any study or the causal inferences drawn from it.”).

1. Temporal relationship.
2. Strength of the association.
3. Dose–response relationship.
4. Replication of the findings.
5. Biological plausibility (coherence with existing knowledge).
6. Consideration of alternative explanations.
7. Cessation of exposure.
8. Specificity of the association, and
9. Consistency with other knowledge.

There is no formula or algorithm that can be used to assess whether a causal inference is appropriate based on these guidelines.¹⁴⁵ One or more factors may be absent even when a true causal relationship exists.¹⁴⁶ Similarly, the existence of some factors does not ensure that a causal relationship exists. Drawing causal inferences after finding an association and considering these factors requires judgment and searching analysis, based on biology, of why a factor or factors may be absent despite a causal relationship, and vice versa. Although the drawing of causal inferences is informed by scientific expertise, it is not a determination that is made by using an objective or algorithmic methodology.

These guidelines reflect criteria proposed by the U.S. Surgeon General in 1964¹⁴⁷ in assessing the relationship between smoking and lung cancer and expanded upon by Sir Austin Bradford Hill in 1965¹⁴⁸ and are often referred to as the Hill criteria or Hill factors.

145. See Douglas L. Weed, *Epidemiologic Evidence and Causal Inference*, 14 *Hematology/Oncology Clinics N. Am.* 797 (2000).

146. See *Cook v. Rockwell Int'l Corp.*, 580 F. Supp. 2d 1071, 1098 (D. Colo. 2006) (rejecting argument that plaintiff failed to provide sufficient evidence of causation based on failing to meet four of the Hill factors).

147. Public Health Serv., U.S. Dep't of Health, Educ., & Welfare, *Smoking and Health: Report of the Advisory Committee to the Surgeon General* (1964); see also Centers for Disease Control and Prevention, U.S. Dep't of Health & Human Servs., *The Health Consequences of Smoking: A Report of the Surgeon General* (2004).

148. See Austin Bradford Hill, *The Environment and Disease: Association or Causation?* 58 *Proc. Royal Soc'y Med.* 295 (1965) (Hill acknowledged that his factors could only serve to assist in the inferential process: "None of my nine viewpoints can bring indisputable evidence for or against the cause-and-effect hypothesis and none can be required as a *sine qua non*."). For discussion of these criteria and their respective strengths in informing a causal inference, see Gordis, *supra* note 32, at 236–39; David E. Lilienfeld & Paul D. Stolley, *Foundations of Epidemiology* 263–66 (3d ed. 1994); Weed, *supra* note 144.

A. Is There a Temporal Relationship?

A temporal, or chronological, relationship must exist for causation to exist. If an exposure causes disease, the exposure must occur before the disease develops.¹⁴⁹ If the exposure occurs after the disease develops, it cannot have caused the disease. Although temporal relationship is often listed as one of many factors in assessing whether an inference of causation is justified, this aspect of a temporal relationship is a necessary factor: Without exposure before the disease, causation cannot exist.¹⁵⁰

With regard to specific causation, a subject dealt with in detail in Section VII, *infra*, there may be circumstances in which a temporal relationship supports the existence of a causal relationship. If the latency period between exposure and outcome is known,¹⁵¹ then exposure consistent with that information may lend credence to a causal relationship. This is particularly true when the latency period is short and competing causes are known and can be ruled out. Thus, if an individual suffers an acute respiratory response shortly after exposure to a suspected agent and other causes of that respiratory problem are known and can be ruled out, the temporal relationship involved supports the conclusion that a causal relationship exists.¹⁵² Similarly, exposure outside a known latency period constitutes evidence, perhaps conclusive evidence, against the existence of causation.¹⁵³ On the other hand, when latency periods are lengthy, variable, or not known and a

149. See *Carroll v. Litton Sys., Inc.*, No. B-C-88-253, 1990 U.S. Dist. LEXIS 16833, at *29 (W.D.N.C. 1990) (“[I]t is essential for . . . [the plaintiffs’ medical experts opining on causation] to know that exposure preceded plaintiffs’ alleged symptoms in order for the exposure to be considered as a possible cause of those symptoms. . . .”).

150. Exposure during the disease initiation process may cause the disease to be more severe than it otherwise would have been without the additional dose.

151. When the latency period is known—or is known to be limited to a specific range of time—as is the case with the adverse effects of some vaccines, the time frame from exposure to manifestation of disease can be critical to determining causation.

152. For courts that have relied on temporal relationships of the sort described, see *Bonner v. ISP Technologies, Inc.*, 259 F.3d 924, 930–31 (8th Cir. 2001) (giving more credence to the expert’s opinion on causation for acute response based on temporal relationship than for chronic disease that plaintiff also developed); *Heller v. Shaw Industries, Inc.* 167 F.3d 146 (3d Cir. 1999); *Westberry v. Gislaved Gummi AB*, 178 F.3d 257 (4th Cir. 1999); *Zuchowicz v. United States*, 140 F.3d 381 (2d Cir. 1998); *Creanga v. Jardal*, 886 A.2d 633, 641 (N.J. 2005); *Alder v. Bayer Corp., AGFA Div.*, 61 P.3d 1068, 1090 (Utah 2002) (“If a bicyclist falls and breaks his arm, causation is assumed without argument because of the temporal relationship between the accident and the injury [and, the court might have added, the absence of any plausible competing causes that might instead be responsible for the broken arm].”).

153. See *In re Phenylpropanolamine (PPA) Prods. Liab. Litig.*, 289 F. Supp. 2d 1230, 1238 (W.D. Wash. 2003) (determining expert testimony on causation for plaintiffs whose exposure was beyond known latency period was inadmissible).

substantial proportion of the disease is due to unknown causes, temporal relationship provides little beyond satisfying the requirement that cause precede effect.¹⁵⁴

*B. How Strong Is the Association Between the Exposure and Disease?*¹⁵⁵

The relative risk is one of the cornerstones for causal inferences.¹⁵⁶ Relative risk measures the strength of the association. The higher the relative risk, the greater the likelihood that the relationship is causal.¹⁵⁷ For cigarette smoking, for example, the estimated relative risk for lung cancer is very high, about 10.¹⁵⁸ That is, the risk of lung cancer in smokers is approximately 10 times the risk in nonsmokers.

A relative risk of 10, as seen with smoking and lung cancer, is so high that it is extremely difficult to imagine any bias or confounding factor that might account for it. The higher the relative risk, the stronger the association and the lower the chance that the effect is spurious. Although lower relative risks can reflect causality, the epidemiologist will scrutinize such associations more closely because there is a greater chance that they are the result of uncontrolled confounding or biases.

154. These distinctions provide a framework for distinguishing between cases that are largely dismissive of temporal relationships as supporting causation and others that find it of significant persuasiveness. Compare cases cited in note 151, *supra*, with *Moore v. Ashland Chem. Inc.*, 151 F.3d 269, 278 (5th Cir. 1998) (giving little weight to temporal relationship in a case in which there were several plausible competing causes that may have been responsible for the plaintiff's disease), and *Glastetter v. Novartis Pharms. Corp.*, 252 F.3d 986, 990 (8th Cir. 2001) (giving little weight to temporal relationship in case studies involving drug and stroke).

155. Assuming that an association is determined to be causal, the strength of the association plays an important role legally in determining the specific causation question—whether the agent caused an individual plaintiff's injury. See *infra* Section VII.

156. See *supra* Section III.A.

157. See *Miller v. Pfizer, Inc.*, 196 F. Supp. 2d 1062, 1079 (D. Kan. 2002) (citing this reference guide); *Landrigan v. Celotex Corp.*, 605 A.2d 1079, 1085 (N.J. 1992). The use of the strength of the association as a factor does not reflect a belief that weaker effects occur less frequently than stronger effects. See *Green*, *supra* note 47, at 652–53 n.39. Indeed, the apparent strength of a given agent is dependent on the prevalence of the other necessary elements that must occur with the agent to produce the disease, rather than on some inherent characteristic of the agent itself. See *Rothman et al.*, *supra* note 61, at 9–11.

158. See *Doll & Hill*, *supra* note 6. The relative risk of lung cancer from smoking is a function of intensity and duration of dose (and perhaps other factors). See *Karen Leffondré et al., Modeling Smoking History: A Comparison of Different Approaches*, 156 *Am. J. Epidemiology* 813 (2002). The relative risk provided in the text is based on a specified magnitude of cigarette exposure.

C. Is There a Dose–Response Relationship?

A dose–response relationship means that the greater the exposure, the greater the risk of disease. Generally, higher exposures should increase the incidence (or severity) of disease.¹⁵⁹ However, some causal agents do not exhibit a dose–response relationship when, for example, there is a threshold phenomenon (i.e., an exposure may not cause disease until the exposure exceeds a certain dose).¹⁶⁰ Thus, a dose–response relationship is strong, but not essential, evidence that the relationship between an agent and disease is causal.¹⁶¹

159. See *Newman v. Motorola, Inc.*, 218 F. Supp. 2d 769, 778 (D. Md. 2002) (recognizing importance of dose–response relationship in assessing causation).

160. The question whether there is a no-effect threshold dose is a controversial one in a variety of toxic substances areas. See, e.g., Irving J. Selikoff, *Disability Compensation for Asbestos-Associated Disease in the United States: Report to the U.S. Department of Labor* 181–220 (1981); Paul Kotin, *Dose–Response Relationships and Threshold Concepts*, 271 Ann. N.Y. Acad. Sci. 22 (1976); K. Robock, *Based on Available Data, Can We Project an Acceptable Standard for Industrial Use of Asbestos? Absolutely*, 330 Ann. N.Y. Acad. Sci. 205 (1979); *Ferebee v. Chevron Chem. Co.*, 736 F.2d 1529, 1536 (D.C. Cir. 1984) (dose–response relationship for low doses is “one of the most sharply contested questions currently being debated in the medical community”); *In re TMI Litig. Consol. Proc.*, 927 F. Supp. 834, 844–45 (M.D. Pa. 1996) (discussing low-dose extrapolation and no-dose effects for radiation exposure).

Moreover, good evidence to support or refute the threshold-dose hypothesis is exceedingly unlikely because of the inability of epidemiology or animal toxicology to ascertain very small effects. Cf. Arnold L. Brown, *The Meaning of Risk Assessment*, 37 *Oncology* 302, 303 (1980). Even the shape of the dose–response curve—whether linear or curvilinear, and if the latter, the shape of the curve—is a matter of hypothesis and speculation. See *Allen v. United States*, 588 F. Supp. 247, 419–24 (D. Utah 1984), *rev'd on other grounds*, 816 F.2d 1417 (10th Cir. 1987); *In re Bextra & Celebrex Mktg. Sales Practices & Prod. Liab. Litig.*, 524 F. Supp. 2d 1166, 1180 (N.D. Cal. 2007) (criticizing expert for “primitive” extrapolation of risk based on assumption of linear relationship of risk to dose); Troyen A. Brennan & Robert F. Carter, *Legal and Scientific Probability of Causation for Cancer and Other Environmental Disease in Individuals*, 10 *J. Health Pol’y & L.* 33, 43–44 (1985).

The idea that the “dose makes the poison” is a central tenet of toxicology and attributed to Paracelsus, in the sixteenth century. See Bernard D. Goldstein & Mary Sue Henfin, *Reference Guide on Toxicology*, Section I.A, in this manual. It does not mean that any agent is capable of causing any disease if an individual is exposed to a sufficient dose. Agents tend to have specific effects, see *infra* Section V.H., and this dictum reflects only the idea that there is a safe dose below which an agent does not cause any toxic effect. See Michael A Gallo, *History and Scope of Toxicology*, in Casarett and Doull’s *Toxicology: The Basic Science of Poisons* 1, 4–5 (Curtis D. Klaassen ed., 7th ed. 2008). For a case in which a party made such a mistaken interpretation of Paracelsus, see *Alder v. Bayer Corp., AGFA Div.*, 61 P.3d 1068, 1088 (Utah 2002). Paracelsus was also responsible for the initial articulation of the specificity tenet. See *infra* Section V.H.

161. Evidence of a dose–response relationship as bearing on whether an inference of general causation is justified is analytically distinct from determining whether evidence of the dose to which a plaintiff was exposed is required in order to establish specific causation. On the latter matter, see *infra* Section VII; *Restatement (Third) of Torts: Liability for Physical and Emotional Harm* § 28 cmt. c(2) & ptrs. note (2010).

D. Have the Results Been Replicated?

Rarely, if ever, does a single study persuasively demonstrate a cause–effect relationship.¹⁶² It is important that a study be replicated in different populations and by different investigators before a causal relationship is accepted by epidemiologists and other scientists.¹⁶³

The need to replicate research findings permeates most fields of science. In epidemiology, research findings often are replicated in different populations.¹⁶⁴ Consistency in these findings is an important factor in making a judgment about causation. Different studies that examine the same exposure–disease relationship generally should yield similar results. Although inconsistent results do not necessarily rule out a causal nexus, any inconsistencies signal a need to explore whether different results can be reconciled with causality.

E. Is the Association Biologically Plausible (Consistent with Existing Knowledge)?¹⁶⁵

Biological plausibility is not an easy criterion to use and depends upon existing knowledge about the mechanisms by which the disease develops. When biological plausibility exists, it lends credence to an inference of causality. For example, the conclusion that high cholesterol is a cause of coronary heart disease is plausible because cholesterol is found in atherosclerotic plaques. However, observations have been made in epidemiologic studies that were not biologically plausible at the time but subsequently were shown to be correct.¹⁶⁶ When an observation is inconsistent with current biological knowledge, it should not be discarded, but

162. In *Kehm v. Procter & Gamble Co.*, 580 F. Supp. 890, 901 (N.D. Iowa 1982), *aff'd*, 724 F.2d 613 (8th Cir. 1983), the court remarked on the persuasive power of multiple independent studies, each of which reached the same finding of an association between toxic shock syndrome and tampon use.

163. This may not be the legal standard, however. *Cf. Smith v. Wyeth-Ayerst Labs. Co.*, 278 F. Supp. 2d 684, 710 n.55 (W.D.N.C. 2003) (observing that replication is difficult to establish when there is only one study that has been performed at the time of trial).

164. *See Cadarian v. Merrell Dow Pharms., Inc.*, 745 F. Supp. 409, 412 (E.D. Mich. 1989) (holding a study on Bendectin insufficient to support an expert's opinion, because "the study's authors themselves concluded that the results could not be interpreted without independent confirmatory evidence").

165. A number of courts have adverted to this criterion in the course of their discussions of causation in toxic substances cases. *E.g., In re Phenylpropanolamine (PPA) Prods. Liab. Litig.*, 289 F. Supp. 2d 1230, 1247–48 (W.D. Wash. 2003); *Cook v. United States*, 545 F. Supp. 306, 314–15 (N.D. Cal. 1982) (discussing biological implausibility of a two-peak increase of disease when plotted against time); *Landrigan v. Celotex Corp.*, 605 A.2d 1079, 1085–86 (N.J. 1992) (discussing the existence vel non of biological plausibility); *see also* Bernard D. Goldstein & Mary Sue Henifin, Reference Guide on Toxicology, Section III.E, in this manual.

166. *See In re Rezulin Prods. Liab. Litig.*, 369 F. Supp. 2d 398, 405 (S.D.N.Y. 2005); *In re Phenylpropanolamine (PPA) Prods. Liab. Litig.*, 289 F. Supp. 2d 1230, 1247 (W.D. Wash. 2003).

the observation should be confirmed before significance is attached to it. The saliency of this factor varies depending on the extent of scientific knowledge about the cellular and subcellular mechanisms through which the disease process works. The mechanisms of some diseases are understood quite well based on the available evidence, including from toxicologic research, whereas other mechanism explanations are merely hypothesized—although hypotheses are sometimes accepted under this factor.¹⁶⁷

F. Have Alternative Explanations Been Considered?

The importance of considering the possibility of bias and confounding and ruling out the possibilities is discussed above.¹⁶⁸

G. What Is the Effect of Ceasing Exposure?

If an agent is a cause of a disease, then one would expect that cessation of exposure to that agent ordinarily would reduce the risk of the disease. This has been the case, for example, with cigarette smoking and lung cancer. In many situations, however, relevant data are simply not available regarding the possible effects of ending the exposure. But when such data are available and eliminating exposure reduces the incidence of disease, this factor strongly supports a causal relationship.

H. Does the Association Exhibit Specificity?

An association exhibits specificity if the exposure is associated only with a single disease or type of disease.¹⁶⁹ The vast majority of agents do not cause a wide vari-

167. See Douglas L. Weed & Stephen D. Hursting, *Biologic Plausibility in Causal Inference: Current Methods and Practice*, 147 *Am. J. Epidemiology* 415 (1998) (examining use of this criterion in contemporary epidemiologic research and distinguishing between alternative explanations of what constitutes biological plausibility, ranging from mere hypotheses to “sufficient evidence to show how the factor influences a known disease mechanism”).

168. See *supra* Sections IV.B–C.

169. This criterion reflects the fact that although an agent causes one disease, it does not necessarily cause other diseases. See, e.g., *Nelson v. Am. Sterilizer Co.*, 566 N.W.2d 671, 676–77 (Mich. Ct. App. 1997) (affirming dismissal of plaintiff’s claims that chemical exposure caused her liver disorder, but recognizing that evidence supported claims for neuropathy and other illnesses); *Sanderson v. Int’l Flavors & Fragrances, Inc.*, 950 F. Supp. 981, 996–98 (C.D. Cal. 1996); see also *Taylor v. Airco, Inc.*, 494 F. Supp. 2d 21, 27 (D. Mass. 2007) (holding that plaintiff’s expert could testify to causal relationship between vinyl chloride and one type of liver cancer for which there was only modest support given strong causal evidence for vinyl chloride and another type of liver cancer).

When a party claims that evidence of a causal relationship between an agent and one disease is relevant to whether the agent caused another disease, courts have required the party to show that

ety of effects. For example, asbestos causes mesothelioma and lung cancer and may cause one or two other cancers, but there is no evidence that it causes any other types of cancers. Thus, a study that finds that an agent is associated with many different diseases should be examined skeptically. Nevertheless, there may be causal relationships in which this guideline is not satisfied. Cigarette manufacturers have long claimed that because cigarettes have been linked to lung cancer, emphysema, bladder cancer, heart disease, pancreatic cancer, and other conditions, there is no specificity and the relationships are not causal. There is, however, at least one good reason why inferences about the health consequences of tobacco do not require specificity: Because tobacco and cigarette smoke are not in fact single agents but consist of numerous harmful agents, smoking represents exposure to multiple agents, with multiple possible effects. Thus, whereas evidence of specificity may strengthen the case for causation, lack of specificity does not necessarily undermine it where there is a good biological explanation for its absence.

I. Are the Findings Consistent with Other Relevant Knowledge?

In addressing the causal relationship of lung cancer to cigarette smoking, researchers examined trends over time for lung cancer and for cigarette sales in the United States. A marked increase in lung cancer death rates in men was observed, which appeared to follow the increase in sales of cigarettes. Had the increase in lung cancer deaths followed a decrease in cigarette sales, it might have given researchers pause. It would not have precluded a causal inference, but the inconsistency of the trends in cigarette sales and lung cancer mortality would have had to be explained.

VI. What Methods Exist for Combining the Results of Multiple Studies?

Not infrequently, the scientific record may include a number of epidemiologic studies whose findings differ. These may be studies in which one shows an association and the other does not, or studies that report associations, but of different

the mechanisms involved in development of the disease are similar. Thus, in *Austin v. Kerr-McGee Refining Corp.*, 25 S.W.3d 280 (Tex. App. 2000), the plaintiff suffered from a specific form of chronic leukemia. Studies demonstrated a causal relationship between benzene and all leukemias, but there was a paucity of evidence on the relationship between benzene and the specific form of leukemia from which plaintiff suffered. The court required that plaintiff's expert demonstrate the similarity of the biological mechanism among leukemias as a condition for the admissibility of his causation testimony, a requirement the court concluded had not been satisfied. *Accord In re Bextra & Celebrex Mktg. Sales Practices & Prod. Liab. Litig.*, 524 F. Supp. 2d 1166, 1183 (N.D. Cal. 2007); *Magistrini v. One Hour Martinizing Dry Cleaning*, 180 F. Supp. 2d 584, 603 (D.N.J. 2002).

magnitude.¹⁷⁰ In view of the fact that studies may disagree and that often many of the studies are small and lack the statistical power needed for definitive conclusions, the technique of meta-analysis was developed, initially for clinical trials.¹⁷¹ Meta-analysis is a method of pooling study results to arrive at a single figure to represent the totality of the studies reviewed.¹⁷² It is a way of systematizing the time-honored approach of reviewing the literature, which is characteristic of science, and placing it in a standardized framework with quantitative methods for estimating risk. In a meta-analysis, studies are given different weights in proportion to the sizes of their study populations and other characteristics.¹⁷³

Meta-analysis is most appropriate when used in pooling randomized experimental trials, because the studies included in the meta-analysis share the most significant methodological characteristics, in particular, use of randomized assignment of subjects to different exposure groups. However, often one is confronted with nonrandomized observational studies of the effects of possible toxic substances or agents. A method for summarizing such studies is greatly needed, but when meta-analysis is applied to observational studies—either case-control or cohort—it becomes more controversial.¹⁷⁴ The reason for this is that often methodological differences among studies are much more pronounced than they are in randomized trials. Hence, the justification for pooling the results and deriving a single estimate of risk, for example, is problematic.¹⁷⁵

170. See, e.g., *Zandi v. Wyeth a/k/a Wyeth, Inc.*, No. 27-CV-06-6744, 2007 WL 3224242 (Minn. Dist. Ct. Oct. 15, 2007) (plaintiff's expert cited 40 studies in support of a causal relationship between hormone therapy and breast cancer; many studies found different magnitudes of increased risk).

171. See *In re Paoli R.R. Yard PCB Litig.*, 916 F.2d 829, 856 (3d Cir. 1990), cert. denied, 499 U.S. 961 (1991); *Hines v. Consol. Rail Corp.*, 926 F.2d 262, 273 (3d Cir. 1991); *Allen v. Int'l Bus. Mach. Corp.*, No. 94-264-LON, 1997 U.S. Dist. LEXIS 8016, at *71–*74 (meta-analysis of observational studies is a controversial subject among epidemiologists). Thus, contrary to the suggestion by at least one court, multiple studies with small numbers of subjects may be pooled to reduce the possibility of sampling error. See *In re Joint E. & S. Dist. Asbestos Litig.*, 827 F. Supp. 1014, 1042 (S.D.N.Y. 1993) (“[N]o matter how many studies yield a positive but statistically insignificant SMR for colorectal cancer, the results remain statistically insignificant. Just as adding a series of zeros together yields yet another zero as the product, adding a series of positive but statistically insignificant SMRs together does not produce a statistically significant pattern.”), rev'd, 52 F.3d 1124 (2d Cir. 1995); see also *supra* note 76.

172. For a nontechnical explanation of meta-analysis, along with case studies of a variety of scientific areas in which it has been employed, see Morton Hunt, *How Science Takes Stock: The Story of Meta-Analysis* (1997).

173. Petitti, *supra* note 88.

174. See Donna F. Stroup et al., *Meta-analysis of Observational Studies in Epidemiology: A Proposal for Reporting*, 283 JAMA 2008, 2009 (2000); Jesse A. Berlin & Carin J. Kim, *The Use of Meta-Analysis in Pharmacoepidemiology*, in *Pharmacoepidemiology* 681, 683–84 (Brian L. Strom ed., 4th ed. 2005).

175. On rare occasions, meta-analyses of both clinical and observational studies are available. See, e.g., *In re Bextra & Celebrex Mktg. Sales Practices & Prod. Liab. Litig.*, 524 F. Supp. 2d 1166, 1175 (N.D. Cal. 2007) (referring to clinical and observational meta-analyses of low dose of a drug; both analyses failed to find any effect).

A number of problems and issues arise in meta-analysis. Should only published papers be included in the meta-analysis, or should any available studies be used, even if they have not been peer reviewed? Can the results of the meta-analysis itself be reproduced by other analysts? When there are several meta-analyses of a given relationship, why do the results of different meta-analyses often disagree? The appeal of a meta-analysis is that it generates a single estimate of risk (along with an associated confidence interval), but this strength can also be a weakness, and may lead to a false sense of security regarding the certainty of the estimate. A key issue is the matter of heterogeneity of results among the studies being summarized. If there is more variance among study results than one would expect by chance, this creates further uncertainty about the summary measure from the meta-analysis. Such differences can arise from variations in study quality, or in study populations or in study designs. Such differences in results make it harder to trust a single estimate of effect; the reasons for such differences need at least to be acknowledged and, if possible, explained.¹⁷⁶ People often tend to have an inordinate belief in the validity of the findings when a single number is attached to them, and many of the difficulties that may arise in conducting a meta-analysis, especially of observational studies such as epidemiologic ones, may consequently be overlooked.¹⁷⁷

VII. What Role Does Epidemiology Play in Proving Specific Causation?

Epidemiology is concerned with the incidence of disease in populations, and epidemiologic studies do not address the question of the cause of an individual's disease.¹⁷⁸ This question, often referred to as specific causation, is beyond the

176. See Stroup et al., *supra* note 173 (recommending methodology for meta-analysis of observational studies).

177. Much has been written about meta-analysis recently and some experts consider the problems of meta-analysis to outweigh the benefits at the present time. For example, John Bailar has observed:

[P]roblems have been so frequent and so deep, and overstatements of the strength of conclusions so extreme, that one might well conclude there is something seriously and fundamentally wrong with the method. For the present . . . I still prefer the thoughtful, old-fashioned review of the literature by a knowledgeable expert who explains and defends the judgments that are presented. We have not yet reached a stage where these judgments can be passed on, even in part, to a formalized process such as meta-analysis.

John C. Bailar III, *Assessing Assessments*, 277 *Science* 528, 529 (1997) (reviewing Morton Hunt, *How Science Takes Stock* (1997)); see also *Point/Counterpoint: Meta-analysis of Observational Studies*, 140 *Am. J. Epidemiology* 770 (1994).

178. See *DeLuca v. Merrell Dow Pharms., Inc.*, 911 F.2d 941, 945 & n.6 (3d Cir. 1990) (“Epidemiological studies do not provide direct evidence that a particular plaintiff was injured by exposure to a substance.”); *In re Viagra Prods. Liab. Litig.*, 572 F. Supp. 2d 1071, 1078 (D. Minn. 2008) (“Epi-

domain of the science of epidemiology. Epidemiology has its limits at the point where an inference is made that the relationship between an agent and a disease is causal (general causation) and where the magnitude of excess risk attributed to the agent has been determined; that is, epidemiologists investigate whether an agent can cause a disease, not whether an agent did cause a specific plaintiff's disease.¹⁷⁹

Nevertheless, the specific causation issue is a necessary legal element in a toxic substance case. The plaintiff must establish not only that the defendant's agent is capable of causing disease, but also that it did cause the plaintiff's disease. Thus, numerous cases have confronted the legal question of what is acceptable proof of specific causation and the role that epidemiologic evidence plays in answering that question.¹⁸⁰ This question is not a question that is addressed by epidemiology.¹⁸¹ Rather, it is a legal question with which numerous courts

demology focuses on the question of general causation (i.e., is the agent capable of causing disease?) rather than that of specific causation (i.e., did it cause a disease in a particular individual?)” (quoting the second edition of this reference guide)); *In re Asbestos Litig.*, 900 A.2d 120, 133 (Del. Super. Ct. 2006); Michael Dore, *A Commentary on the Use of Epidemiological Evidence in Demonstrating Cause-in-Fact*, 7 Harv. Envtl. L. Rev. 429, 436 (1983).

There are some diseases that do not occur without exposure to a given toxic agent. This is the same as saying that the toxic agent is a necessary cause for the disease, and the disease is sometimes referred to as a signature disease (also, the agent is pathognomonic), because the existence of the disease necessarily implies the causal role of the agent. See Kenneth S. Abraham & Richard A. Merrill, *Scientific Uncertainty in the Courts*, Issues Sci. & Tech. 93, 101 (1986). Asbestosis is a signature disease for asbestos, and vaginal adenocarcinoma (in young adult women) is a signature disease for in utero DES exposure.

179. *Cf. In re “Agent Orange” Prod. Liab. Litig.*, 597 F. Supp. 740, 780 (E.D.N.Y. 1984) (Agent Orange allegedly caused a wide variety of diseases in Vietnam veterans and their offspring), *aff’d*, 818 F.2d 145 (2d Cir. 1987).

180. In many instances, causation can be established without epidemiologic evidence. When the mechanism of causation is well understood, the causal relationship is well established, or the timing between cause and effect is close, scientific evidence of causation may not be required. This is frequently the situation when the plaintiff suffers traumatic injury rather than disease. This section addresses only those situations in which causation is not evident, and scientific evidence is required.

181. Nevertheless, an epidemiologist may be helpful to the factfinder in answering this question. Some courts have permitted epidemiologists (or those who use epidemiologic methods) to testify about specific causation. See *Ambrosini v. Labarraque*, 101 F.3d 129, 137–41 (D.C. Cir. 1996); *Zuchowicz v. United States*, 870 F. Supp. 15 (D. Conn. 1994); *Landrigan v. Celotex Corp.*, 605 A.2d 1079, 1088–89 (N.J. 1992). In general, courts seem more concerned with the basis of an expert’s opinion than with whether the expert is an epidemiologist or clinical physician. See *Porter v. Whitehall*, 9 F.3d 607, 614 (7th Cir. 1992) (“curb side” opinion from clinician not admissible); *Burton v. R.J. Reynolds Tobacco Co.*, 181 F. Supp. 2d 1256, 1266–67 (D. Kan. 2002) (vascular surgeon permitted to testify to general causation over objection based on fact he was not an epidemiologist); *Wade-Greaux v. Whitehall Labs.*, 874 F. Supp. 1441, 1469–72 (D.V.I.) (clinician’s multiple bases for opinion inadequate to support causation opinion), *aff’d*, 46 F.3d 1120 (3d Cir. 1994); *Landrigan*, 605 A.2d at 1083–89 (permitting both clinicians and epidemiologists to testify to specific causation provided the methodology used is sound); *Trach v. Fellin*, 817 A.2d 1102, 1118–19 (Pa. Super. Ct. 2003) (toxicologist and pathologist permitted to testify to specific causation).

have grappled.¹⁸² The remainder of this section is predominantly an explanation of judicial opinions. It is, in addition, in its discussion of the reasoning behind applying the risk estimates of an epidemiologic body of evidence to an individual, informed by epidemiologic principles and methodological research.

Before proceeding, one more caveat is in order. This section assumes that epidemiologic evidence has been used as proof of causation for a given plaintiff. The discussion does not address whether a plaintiff must use epidemiologic evidence to prove causation.¹⁸³

Two legal issues arise with regard to the role of epidemiology in proving individual causation: admissibility and sufficiency of evidence to meet the burden of production. The first issue tends to receive less attention by the courts but nevertheless deserves mention. An epidemiologic study that is sufficiently rigorous to justify a conclusion that it is scientifically valid should be admissible,¹⁸⁴ as it tends to make an issue in dispute more or less likely.¹⁸⁵

182. See Restatement (Third) of Torts: Liability for Physical and Emotional Harm § 28 cmt. c(3) (2010) (“Scientists who conduct group studies do not examine specific causation in their research. No scientific methodology exists for assessing specific causation for an individual based on group studies. Nevertheless, courts have reasoned from the preponderance-of-the-evidence standard to determine the sufficiency of scientific evidence on specific causation when group-based studies are involved”).

183. See *id.* § 28 cmt. c(3) & rptrs. note (“most courts have appropriately declined to impose a threshold requirement that a plaintiff always must prove causation with epidemiologic evidence”); see also *Westberry v. Gislaved Gummi AB*, 178 F.2d 257 (4th Cir. 1999) (acute response, differential diagnosis ruled out other known causes of disease, dechallenge, rechallenge tests by expert that were consistent with exposure to defendant’s agent causing disease, and absence of epidemiologic or toxicologic studies; holding that expert’s testimony on causation was properly admitted); *Zuchowicz v. United States*, 140 F.3d 381 (2d Cir. 1998); *In re Heparin Prods. Liab. Litig.* 2011 WL 2971918, at *7-10 (N.D. Ohio July 21, 2011).

184. See *DeLuca v. Merrell Dow Pharms., Inc.*, 911 F.2d 941, 958 (3d Cir. 1990); *cf.* *Kehm v. Procter & Gamble Co.*, 580 F. Supp. 890, 902 (N.D. Iowa 1982) (“These [epidemiologic] studies were highly probative on the issue of causation—they all concluded that an association between tampon use and menstrually related TSS [toxic shock syndrome] cases exists.”), *aff’d*, 724 F.2d 613 (8th Cir. 1984).

Hearsay concerns may limit the independent admissibility of the study, but the study could be relied on by an expert in forming an opinion and may be admissible pursuant to Fed. R. Evid. 703 as part of the underlying facts or data relied on by the expert.

In *Ellis v. International Playtex, Inc.*, 745 F.2d 292, 303 (4th Cir. 1984), the court concluded that certain epidemiologic studies were admissible despite criticism of the methodology used in the studies. The court held that the claims of bias went to the studies’ weight rather than their admissibility. *Cf.* *Christophersen v. Allied-Signal Corp.*, 939 F.2d 1106, 1109 (5th Cir. 1991) (“As a general rule, questions relating to the bases and sources of an expert’s opinion affect the weight to be assigned that opinion rather than its admissibility. . . .”).

185. Even if evidence is relevant, it may be excluded if its probative value is substantially outweighed by prejudice, confusion, or inefficiency. Fed. R. Evid. 403. However, exclusion of an otherwise relevant epidemiologic study on Rule 403 grounds is unlikely.

In *Daubert v. Merrell Dow Pharmaceuticals, Inc.*, 509 U.S. 579, 591 (1993), the Court invoked the concept of “fit,” which addresses the relationship of an expert’s scientific opinion to the facts of the case and the issues in dispute. In a toxic substance case in which cause in fact is disputed, an epi-

Far more courts have confronted the role that epidemiology plays with regard to the sufficiency of the evidence and the burden of production.¹⁸⁶ The civil burden of proof is described most often as requiring belief by the factfinder “that what is sought to be proved is more likely true than not true.”¹⁸⁷ The relative risk from epidemiologic studies can be adapted to this 50%-plus standard to yield a probability or likelihood that an agent caused an individual’s disease.¹⁸⁸ An important caveat is necessary, however. The discussion below speaks in terms of the magnitude of the relative risk or association found in a study. However, before an association or relative risk is used to make a statement about the probability of individual causation, the inferential judgment, described in Section V, that the association is truly causal rather than spurious, is required: “[A]n agent cannot be considered to cause the illness of a specific person unless it is recognized as a cause of that disease in general.”¹⁸⁹ The following discussion should be read with this caveat in mind.¹⁹⁰

demographic study of the same agent to which the plaintiff was exposed that examined the association with the same disease from which the plaintiff suffers would undoubtedly have sufficient “fit” to be a part of the basis of an expert’s opinion. The Court’s concept of “fit,” borrowed from *United States v. Downing*, 753 F.2d 1224, 1242 (3d Cir. 1985), appears equivalent to the more familiar evidentiary concept of probative value, albeit one requiring assessment of the scientific reasoning the expert used in drawing inferences from methodology or data to opinion.

186. We reiterate a point made at the outset of this section: This discussion of the use of a threshold relative risk for specific causation is not epidemiology or an inquiry an epidemiologist would undertake. This is an effort by courts and commentators to adapt the legal standard of proof to the available scientific evidence. See *supra* text accompanying notes 175–179. While strength of association is a guideline for drawing an inference of causation from an association, see *supra* Section V, there is no specified threshold required.

187. Kevin F. O’Malley et al., *Federal Jury Practice and Instructions* § 104.01 (5th ed. 2000); see also *United States v. Fatico*, 458 F. Supp. 388, 403 (E.D.N.Y. 1978) (“Quantified, the preponderance standard would be 50%+ probable.”), *aff’d*, 603 F.2d 1053 (2d Cir. 1979).

188. An adherent of the frequentist school of statistics would resist this adaptation, which may explain why many epidemiologists and toxicologists also resist it. To take the step identified in the text of using an epidemiologic study outcome to determine the probability of specific causation requires a shift from a frequentist approach, which involves sampling or frequency data from an empirical test, to a subjective probability about a discrete event. Thus, a frequentist might assert, after conducting a sampling test, that 60% of the balls in an opaque container are blue. The same frequentist would resist the statement, “The probability that a single ball removed from the box and hidden behind a screen is blue is 60%.” The ball is either blue or not, and no frequentist data would permit the latter statement. “[T]here is no logically rigorous definition of what a statement of probability means with reference to an individual instance. . . .” Lee Loevinger, *On Logic and Sociology*, 32 *Jurimetrics J.* 527, 530 (1992); see also Steve Gold, *Causation in Toxic Torts: Burdens of Proof, Standards of Persuasion and Statistical Evidence*, 96 *Yale L.J.* 376, 382–92 (1986). Subjective probabilities about unique events are employed by those using Bayesian methodology. See Kaye, *supra* note 80, at 54–62; David H. Kaye & David A. Freedman, *Reference Guide on Statistics*, Section IV.D, in this manual.

189. Cole, *supra* note 65, at 10,284.

190. We emphasize this caveat, both because it is not intuitive and because some courts have failed to appreciate the difference between an association and a causal relationship. See, e.g., *Forsyth v. Eli Lilly & Co.*, Civ. No. 95-00185 ACK, 1998 U.S. Dist. LEXIS 541, at *26–*31 (D. Haw. Jan. 5, 1998). *But see*

Some courts have reasoned that when epidemiologic studies find that exposure to the agent causes an incidence in the exposed group that is more than twice the incidence in the unexposed group (i.e., a relative risk greater than 2.0), the probability that exposure to the agent caused a similarly situated individual's disease is greater than 50%.¹⁹¹ These courts, accordingly, hold that when there is group-based evidence finding that exposure to an agent causes an incidence of disease in the exposed group that is more than twice the incidence in the unexposed group, the evidence is sufficient to satisfy the plaintiff's burden of production and permit submission of specific causation to a jury. In such a case, the factfinder may find that it is more likely than not that the substance caused the particular plaintiff's disease. Courts, thus, have permitted expert witnesses to testify to specific causation based on the logic of the effect of a doubling of the risk.¹⁹²

While this reasoning has a certain logic as far as it goes, there are a number of significant assumptions and important caveats that require explication:

1. A valid study and risk estimate. The propriety of this "doubling" reasoning depends on group studies identifying a genuine causal relationship and a reasonably reliable measure of the increased risk.¹⁹³ This requires attention

Berry v. CSX Transp., Inc., 709 So. 2d 552, 568 (Fla. Dist. Ct. App. 1998) ("From epidemiologic studies demonstrating an association, an epidemiologist may or may not infer that a causal relationship exists.").

191. An alternative, yet similar, means to address probabilities in individual cases is use of the attributable fraction parameter, also known as the attributable risk. *See supra* Section III.C. The attributable fraction is that portion of the excess risk that can be attributed to an agent, above and beyond the background risk that is due to other causes. Thus, when the relative risk is greater than 2.0, the attributable fraction exceeds 50%.

192. For a comprehensive list of cases that support proof of causation based on group studies, see Restatement (Third) of Torts: Liability for Physical and Emotional Harm § 28 cmt. c(4) rptrs. note (2010). The Restatement catalogues those courts that require a relative risk in excess of 2.0 as a threshold for sufficient proof of specific causation and those courts that recognize that a lower relative risk than 2.0 can support specific causation, as explained below. Despite considerable disagreement on whether a relative risk of 2.0 is required or merely a taking-off point for determining the sufficiency of the evidence on specific causation, two commentators who surveyed the cases observed that "[t] here were no clear differences in outcomes as between federal and state courts." Russel Lynn S. Carruth & Bernard D. Goldstein, *Relative Risk Greater than Two in Proof of Causation in Toxic Tort Litigation*, 41 *Jurimetrics J.* 195, 199 (2001).

193. Indeed, one commentator contends that, because epidemiology is sufficiently imprecise to accurately measure small increases in risk, in general, studies that find a relative risk less than 2.0 should not be sufficient for causation. The concern is not with specific causation but with general causation and the likelihood that an association less than 2.0 is noise rather than reflecting a true causal relationship. *See* Michael D. Green, *The Future of Proportional Liability*, in *Exploring Tort Law* (Stuart Madden ed., 2005); *see also* Samuel M. Lesko & Allen A. Mitchell, *The Use of Randomized Controlled Trials for Pharmacoepidemiology Studies*, in *Pharmacoepidemiology* 599, 601 (Brian L. Strom ed., 4th ed. 2005) ("it is advisable to use extreme caution in making causal inferences from small relative risks derived from observational studies"); Gary Taubes, *Epidemiology Faces Its Limits*, 269 *Science* 164 (1995) (explaining views of several epidemiologists about a threshold relative risk of 3.0 to seriously consider a causal relationship); N.E. Breslow & N.E. Day, *Statistical Methods in Cancer Research*, in *The Analysis*

to the possibility of random error, bias, or confounding being the source of the association rather than a true causal relationship as explained in Sections IV and V, supra.¹⁹⁴

2. Similarity among study subjects and plaintiff. Only if the study subjects and the plaintiff are similar with respect to other risk factors will a risk estimate from a study or studies be valid when applied to an individual.¹⁹⁵ Thus, if those exposed in a study of the risk of lung cancer from smoking smoked half a pack of cigarettes a day for 20 years, the degree of increased incidence of lung cancer among them cannot be extrapolated to someone who smoked two packs of cigarettes for 30 years without strong (and questionable) assumptions about the dose–response relationship.¹⁹⁶ This is also applicable to risk factors for competing causes. Thus, if all of the subjects in a study are participating because they were identified as having a family history of heart disease, the magnitude of risk found in a study of smok-

of Case-Control Studies 36 (IARC Pub. No. 32, 1980) (“[r]elative risks of less than 2.0 may readily reflect some unperceived bias or confounding factor”); David A. Freedman & Philip B. Stark, *The Swine Flu Vaccine and Guillain-Barré Syndrome: A Case Study in Relative Risk and Specific Causation*, 64 *Law & Contemp. Probs.* 49, 61 (2001) (“If the relative risk is near 2.0, problems of bias and confounding in the underlying epidemiologic studies may be serious, perhaps intractable.”).

194. An excellent explanation for why differential diagnoses generally are inadequate without further proof of general causation was provided in *Cavallo v. Star Enterprises*, 892 F. Supp. 756 (E.D. Va. 1995), *aff’d in relevant part*, 100 F.3d 1150 (4th Cir. 1996):

The process of differential diagnosis is undoubtedly important to the question of “specific causation”. If other possible causes of an injury cannot be ruled out, or at least the probability of their contribution to causation minimized, then the “more likely than not” threshold for proving causation may not be met. But, it is also important to recognize that a fundamental assumption underlying this method is that the final, suspected “cause” remaining after this process of elimination must actually be capable of causing the injury. That is, the expert must “rule in” the suspected cause as well as “rule out” other possible causes. And, of course, expert opinion on this issue of “general causation” must be derived from a scientifically valid methodology.

Id. at 771 (footnote omitted); *see also* *Ruggiero v. Warner-Lambert Co.*, 424 F.3d 249, 254 (2d Cir. 2005); *Norris v. Baxter Healthcare Corp.*, 397 F.3d 878, 885 (10th Cir. 2005); *Meister v. Med. Eng’g Corp.*, 267 F.3d 1123, 1128–29 (D.C. Cir. 2001); *Bickel v. Pfizer, Inc.*, 431 F. Supp. 2d 918, 923–24 (N.D. Ind. 2006); *In re Rezulin Prods. Liab. Litig.*, 369 F. Supp. 2d 398, 436 (S.D.N.Y. 2005); *Coastal Tankships, U.S.A., Inc. v. Anderson*, 87 S.W.3d 591, 608–09 (Tex. Ct. App. 2002); *see generally* Joseph Sanders & Julie Machal-Fulks, *The Admissibility of Differential Diagnosis Testimony to Prove Causation in Toxic Tort Cases: The Interplay of Adjective and Substantive Law*, 64 *Law & Contemp. Probs.* 107, 122–25 (2001) (discussing cases rejecting differential diagnoses in the absence of other proof of general causation and contrary cases).

195 “The basic premise of probability of causation is that individual risk can be determined from epidemiologic data for a representative population; however the premise only holds if the individual is truly representative of the reference population.” Council on Scientific Affairs, *American Medical Association, Radioepidemiological Tables* 257 JAMA 806 (1987).

196. Conversely, a risk estimate from a study that involved a greater exposure is not applicable to an individual exposed to a lower dose. *See, e.g., In re Bextra & Celebrex Mktg. Sales Practices & Prod. Liab. Litig.*, 524 F. Supp. 2d 1166, 1175–76 (N.D. Cal. 2007) (relative risk found in studies of those who took twice the dose of others could not support expert’s opinion of causation for latter group).

ing on the risk of heart disease cannot validly be applied to an individual without such a family history. Finally, if an individual has been differentially exposed to other risk factors from those in a study, the results of the study will not provide an accurate basis for the probability of causation for the individual.¹⁹⁷ Consider once again a study of the effect of smoking on lung cancer among subjects who have no asbestos exposure. The relative risk of smoking in that study would not be applicable to an asbestos insulation worker. More generally, if the study subjects are heterogeneous with regard to risk factors related to the outcome of interest, the relative risk found in a study represents *an average risk for the group* rather than a uniform increased risk applicable to each individual.¹⁹⁸

3. *Nonacceleration of disease.* Another assumption embedded in using the risk findings of a group study to determine the probability of causation in an individual is that the disease is one that never would have been contracted absent exposure. Put another way, the assumption is that the agent did not merely accelerate occurrence of the disease without affecting the lifetime risk of contracting the disease. Birth defects are an example of an outcome that is not accelerated. However, for most of the chronic diseases of adulthood, it is not possible for epidemiologic studies to distinguish between acceleration of disease and causation of new disease. If, in fact, acceleration

197. See David H. Kaye & David A. Freedman, Reference Guide on Statistics, in this manual (explaining the problems of employing a study outcome to determine the probability of an individual's having contracted the disease from exposure to the agent because of variations in individuals that bear on the risk of a given individual contracting the disease); David A. Freedman & Philip Stark, *The Swine Flu Vaccine and Guillain-Barré Syndrome: A Case Study in Relative Risk and Specific Causation*, 23 Evaluation Rev. 619 (1999) (analyzing the role that individual variation plays in determining the probability of specific causation based on the relative risk found in a study and providing a mathematical model for calculating the effect of individual variation); Mark Parascandola, *What Is Wrong with the Probability of Causation?* 39 Jurimetrics J. 29 (1998).

198. The comment of two prominent epidemiologists on this subject is illuminating:

We cannot measure the individual risk, and assigning the average value to everyone in the category reflects nothing more than our ignorance about the determinants of lung cancer that interact with cigarette smoke. It is apparent from epidemiological data that some people can engage in chain smoking for many decades without developing lung cancer. Others are or will become primed by unknown circumstances and need only to add cigarette smoke to the nearly sufficient constellation of causes to initiate lung cancer. In our ignorance of these hidden causal components, the best we can do in assessing risk is to classify people according to measured causal risk indicators and then assign the average observed within a class to persons within the class.

Rothman & Greenland, *supra* note 131, at 9; see also Ofer Shpilberg et al., *The Next Stage: Molecular Epidemiology*, 50 J. Clinical Epidemiology 633, 637 (1997) (“A 1.5-fold relative risk may be composed of a 5-fold risk in 10% of the population, and a 1.1-fold risk in the remaining 90%, or a 2-fold risk in 25% and a 1.1-fold for 75%, or a 1.5-fold risk for the entire population.”).

- is involved, the relative risk from a study will understate the probability that exposure accelerated the occurrence of the disease.¹⁹⁹
4. *Agent operates independently.* Employing a risk estimate to determine the probability of causation is not valid if the agent interacts with another cause in a way that results in an increase in disease beyond merely the sum of the increased incidence due to each agent separately. For example, the relative risk of lung cancer due to smoking is around 10, while the relative risk for asbestos exposure is approximately 5. The relative risk for someone exposed to both is not the arithmetic sum of the two relative risks, that is, 15, but closer to the product (50- to 60-fold), reflecting an interaction between the two.²⁰⁰ Neither of the individual agent's relative risks can be employed to estimate the probability of causation in someone exposed to both asbestos and cigarette smoke.²⁰¹
 5. *Other assumptions.* Additional assumptions include (a) the agent of interest is not responsible for fatal diseases other than the disease of interest²⁰² and (b) the agent does not provide a protective effect against the outcome of interest in a subpopulation of those being studied.²⁰³

Evidence in a given case may challenge one or more of these assumptions. Bias in a study may suggest that the study findings are inaccurate and should be estimated to be higher or lower or, even, that the findings are spurious, that is, do not reflect a true causal relationship. A plaintiff may have been exposed to a

199. See Sander Greenland & James M. Robins, *Epidemiology, Justice, and the Probability of Causation*, 40 *Jurimetrics J.* 321 (2000); Sander Greenland, *Relation of Probability of Causation to Relative Risk and Doubling Dose: A Methodologic Error That Has Become a Social Problem*, 89 *Am. J. Pub. Health* 1166 (1999). If acceleration occurs, then the appropriate characterization of the harm for purposes of determining damages would have to be addressed. A defendant who only accelerates the occurrence of harm, say, chronic back pain, that would have occurred independently in the plaintiff at a later time is not liable for the same amount of damages as a defendant who causes a lifetime of chronic back pain. See David A. Fischer, *Successive Causes and the Enigma of Duplicated Harm*, 66 *Tenn. L. Rev.* 1127, 1127 (1999); Michael D. Green, *The Intersection of Factual Causation and Damages*, 55 *DePaul L. Rev.* 671 (2006).

200. We use interaction to mean that the combined effect is other than the additive sum of each effect, which is what we would expect if the two agents operate independently. Statisticians employ the term interaction in a different manner to mean the outcome deviates from what was expected in the model specified in advance. See Jay S. Kaufman, *Interaction Reaction*, 20 *Epidemiology* 159 (2009); Sander Greenland & Kenneth J. Rothman, *Concepts of Interaction*, in Rothman & Greenland, *supra* note 131, at 329.

201. See Restatement (Third) of Torts: Liability for Physical and Emotional Harm § 28 cmt. c(5) (2010); Jan Beyea & Sander Greenland, *The Importance of Specifying the Underlying Biologic Model in Estimating the Probability of Causation*, 76 *Health Physics* 269 (1999).

202. This is because in the epidemiologic studies relied on, those deaths caused by the alternative disease process will mask the true magnitude of increased incidence of the studied disease when the study subjects die before developing the disease of interest.

203. See Greenland & Robins, *supra*, note 198, at 332–33.

dose of the agent in question that is greater or lower than that to which those in the study were exposed.²⁰⁴ A plaintiff may have individual factors, such as higher age than those in the study, that make it less likely that exposure to the agent caused the plaintiff's disease. Similarly, an individual plaintiff may be able to rule out other known (background) causes of the disease, such as genetics, that increase the likelihood that the agent was responsible for that plaintiff's disease. Evidence of a pathological mechanism may be available for the plaintiff that is relevant to the cause of the plaintiff's disease.²⁰⁵ Before any causal relative risk from an epidemiologic study can be used to estimate the probability that the agent in question caused an individual plaintiff's disease, consideration of these (and related) factors is required.²⁰⁶

Having additional evidence that bears on individual causation has led a few courts to conclude that a plaintiff may satisfy his or her burden of production even if a relative risk less than 2.0 emerges from the epidemiologic evidence.²⁰⁷ For example, genetics might be known to be responsible for 50% of the incidence of a disease independent of exposure to the agent.²⁰⁸ If genetics can be ruled out

204. See *supra* Section V.C; see also *Ferebee v. Chevron Chem. Co.*, 736 F.2d 1529, 1536 (D.C. Cir. 1984) ("The dose–response relationship at low levels of exposure for admittedly toxic chemicals like paraquat is one of the most sharply contested questions currently being debated in the medical community."); *In re Joint E. & S. Dist. Asbestos Litig.*, 774 F. Supp. 113, 115 (S.D.N.Y. 1991) (discussing different relative risks associated with different doses), *rev'd on other grounds*, 964 F.2d 92 (2d Cir. 1992).

205. See *Tobin v. Astra Pharm. Prods., Inc.*, 993 F.2d 528 (6th Cir. 1993) (plaintiff's expert relied predominantly on pathogenic evidence).

206. See *Merrell Dow Pharms., Inc. v. Havner*, 953 S.W.2d 706, 720 (Tex. 1997); *Smith v. Wyeth-Ayerst Labs. Co.*, 278 F. Supp. 2d 684, 708–09 (W.D.N.C. 2003) (describing expert's effort to refine relative risk applicable to plaintiff based on specific risk characteristics applicable to her, albeit in an ill-explained manner); *McDarby v. Merck & Co.*, 949 A.2d 223 (N.J. Super. Ct. App. Div. 2008); *Mary Carter Andruess, Proof of Cancer Causation in Toxic Waste Litigation*, 61 S. Cal. L. Rev. 2075, 2100–04 (1988). An example of a judge sitting as factfinder and considering individual factors for a number of plaintiffs in deciding cause in fact is contained in *Allen v. United States*, 588 F. Supp. 247, 429–43 (D. Utah 1984), *rev'd on other grounds*, 816 F.2d 1417 (10th Cir. 1987), *cert. denied*, 484 U.S. 1004 (1988); see also *Manko v. United States*, 636 F. Supp. 1419, 1437 (W.D. Mo. 1986), *aff'd*, 830 F.2d 831 (8th Cir. 1987).

207. *In re Hanford Nuclear Reservation Litig.*, 292 F.3d 1124, 1137 (9th Cir. 2002) (applying Washington law) (recognizing the role of individual factors that may modify the probability of causation based on the relative risk); *Magistrini v. One Hour Martinizing Dry Cleaning*, 180 F. Supp. 2d 584, 606 (D.N.J. 2002) ("[A] relative risk of 2.0 is not so much a password to a finding of causation as one piece of evidence, among others for the court to consider in determining whether an expert has employed a sound methodology in reaching his or her conclusion."); *Miller v. Pfizer, Inc.*, 196 F. Supp. 2d 1062, 1079 (D. Kan. 2002) (rejecting a threshold of 2.0 for the relative risk and recognizing that even a relative risk greater than 2.0 may be insufficient); *Pafford v. Sec'y, Dept. of Health & Human Servs.*, 64 Fed. Cl. 19 (2005) (acknowledging that epidemiologic studies finding a relative risk of less than 2.0 can provide supporting evidence of causation), *aff'd*, 451 F.3d 1352 (Fed. Cir. 2006).

208. See generally *Steve C. Gold, The More We Know, the Less Intelligent We Are? How Genomic Information Should, and Should Not, Change Toxic Tort Causation Doctrine*, 34 *Harv. Envtl. L. Rev.* 369 (2010); *Jamie A. Grodsky, Genomics and Toxic Torts: Dismantling the Risk-Injury Divide*, 59 *Stan. L. Rev.*

in an individual's case, then a relative risk greater than 1.5 might be sufficient to support an inference that the agent was more likely than not responsible for the plaintiff's disease.²⁰⁹

Indeed, this idea of eliminating a known and competing cause is central to the methodology popularly known in legal terminology as differential diagnosis²¹⁰ but is more accurately referred to as differential etiology.²¹¹ Nevertheless, the logic is sound if the label is not: Eliminating other known and competing causes increases the probability that a given individual's disease was caused by exposure to the agent. In a differential etiology, an expert first determines other known causes of the disease in question and then attempts to ascertain whether those competing causes can be "ruled out" as a cause of plaintiff's disease²¹² as in the

1671 (2007); Gary E. Marchant, *Genetic Data in Toxic Tort Litigation*, 14 J.L. & Pol'y 7 (2006); Gary E. Marchant, *Genetics and Toxic Torts*, 31 Seton Hall L. Rev. 949 (2001).

209. The use of probabilities in excess of .50 to support a verdict results in an all-or-nothing approach to damages that some commentators have criticized. The criticism reflects the fact that defendants responsible for toxic agents with a relative risk just above 2.0 may be required to pay damages not only for the disease that their agents caused, but also for all instances of the disease. Similarly, those defendants whose agents increase the risk of disease by less than a doubling may not be required to pay damages for any of the disease that their agents caused. See, e.g., 2 American Law Inst., Reporter's Study on Enterprise Responsibility for Personal Injury: Approaches to Legal and Institutional Change 369-75 (1991). Judge Posner has been in the vanguard of those advocating that damages be awarded on a proportional basis that reflects the probability of causation or liability. See, e.g., *Doll v. Brown*, 75 F.3d 1200, 1206-07 (7th Cir. 1996). To date, courts have not adopted a rule that would apportion damages based on the probability of cause in fact in toxic substances cases. See Green, *supra* note 192.

210. Physicians regularly employ differential diagnoses in treating their patients to identify the disease from which the patient is suffering. See Jennifer R. Jamison, *Differential Diagnosis for Primary Practice* (1999).

211. It is important to emphasize that the term "differential diagnosis" in a clinical context refers to identifying a set of diseases or illnesses responsible for the patient's symptoms, while "differential etiology" refers to identifying the causal factors involved in an individual's disease or illness. For many health conditions, the *cause* of the disease or illness has no relevance to its treatment, and physicians, therefore, do not employ this term or pursue that question. See *Zandi v. Wyeth a/k/a Wyeth, Inc.*, No. 27-CV-06-6744, 2007 WL 3224242 (Minn. Dist. Ct. Oct. 15, 2007) (commenting that physicians do not attempt to determine the cause of breast cancer). Thus, the standard differential diagnosis performed by a physician is not to determine the cause of a patient's disease. See John B. Wong et al., Reference Guide on Medical Testimony, in this manual; Edward J. Imwinkelried, *The Admissibility and Legal Sufficiency of Testimony About Differential Diagnosis (Etiology): of Under — and Over — Estimations*, 56 Baylor L. Rev. 391, 402-03 (2004); see also *Turner v. Iowa Fire Equip. Co.*, 229 F.3d 1202, 1208 (8th Cir. 2000) (distinguishing between differential diagnosis conducted for the purpose of identifying the disease from which the patient suffers and one attempting to determine the cause of the disease); *Creanga v. Jardal*, 886 A.2d 633, 639 (N.J. 2005) ("Whereas most physicians use the term to describe the process of determining which of several diseases is causing a patient's *symptoms*, courts have used the term in a more general sense to describe the process by which causes of the patient's *condition* are identified.").

212. Courts regularly affirm the legitimacy of employing differential diagnostic methodology. See, e.g., *In re Ephedra Prods. Liab. Litig.*, 393 F. Supp. 2d 181, 187 (S.D.N.Y. 2005); *Easum v. Miller*, 92 P.3d 794, 802 (Wyo. 2004) ("Most circuits have held that a reliable differential diagnosis satisfies *Daubert* and provides a valid foundation for admitting an expert opinion. The circuits reason that a differential diagnosis is a tested methodology, has been subjected to peer review/publication, does not

genetics example in the preceding paragraph. Similarly, an expert attempting to determine whether an individual's emphysema was caused by occupational chemical exposure would inquire whether the individual was a smoker. By ruling out (or ruling in) the possibility of other causes, the probability that a given agent was the cause of an individual's disease can be refined. Differential etiologies are most critical when the agent at issue is relatively weak and is not responsible for a large proportion of the disease in question.

Although differential etiologies are a sound methodology in principle, this approach is only valid if general causation exists and a substantial proportion of competing causes are known.²¹³ Thus, for diseases for which the causes are largely unknown, such as most birth defects, a differential etiology is of little benefit.²¹⁴ And, like any scientific methodology, it can be performed in an unreliable manner.²¹⁵

VIII. Acknowledgments

The authors are grateful for the able research assistance provided by Murphy Horne, Wake Forest Law School class of 2012, and Cory Randolph, Wake Forest Law School class of 2010.

frequently lead to incorrect results, and is generally accepted in the medical community.” (quoting *Turner v. Iowa Fire Equip. Co.*, 229 F.3d 1202, 1208 (8th Cir. 2000)); *Alder v. Bayer Corp.*, AGFA Div., 61 P.3d 1068, 1084–85 (Utah 2002).

213. Courts have long recognized that to prove causation plaintiff need not eliminate *all* potential competing causes. *See Stubbs v. City of Rochester*, 134 N.E. 137, 140 (N.Y. 1919) (rejecting defendant's argument that plaintiff was required to eliminate all potential competing causes of typhoid); *see also Easum v. Miller*, 92 P.3d 794, 804 (Wyo. 2004). At the same time, before a competing cause should be considered relevant to a differential diagnosis, there must be adequate evidence that it *is* a cause of the disease. *See Cooper v. Smith & Nephew, Inc.*, 259 F.3d 194, 202 (4th Cir. 2001); *Ranes v. Adams Labs., Inc.*, 778 N.W.2d 677, 690 (Iowa 2010).

214. *See Perry v. Novartis Pharms. Corp.*, 564 F. Supp. 2d 452, 469 (E.D. Pa. 2008) (finding experts' testimony inadmissible because of failure to account for idiopathic (unknown) causes in conducting differential diagnosis); *Soldo v. Sandoz Pharms. Corp.*, 244 F. Supp. 2d 434, 480, 519 (W.D. Pa. 2003) (criticizing expert for failing to account for idiopathic causes); *Magistrini v. One Hour Martinizing Dry Cleaning*, 180 F. Supp. 2d 584, 609 (D.N.J. 2002) (observing that 90–95% of leukemias are of unknown causes, but proceeding incorrectly to assert that plaintiff was obliged to prove that her exposure to defendant's benzene was *the* cause of her leukemia rather than simply a cause of the disease that combined with other exposures to benzene). *But see Ruff v. Ensign-Bickford Indus., Inc.*, 168 F. Supp. 2d 1271, 1286 (D. Utah 2001) (responding to defendant's evidence that most instances of disease are of unknown origin by stating that such matter went to the weight to be attributed to plaintiff's expert's testimony not its admissibility).

215. Numerous courts have concluded that, based on the manner in which a differential diagnosis was conducted, it was unreliable and the expert's testimony based on it is inadmissible. *See, e.g., Glastetter v. Novartis Pharms. Corp.*, 252 F.3d 986, 989 (8th Cir. 2001).

Glossary of Terms

The following terms and definitions were adapted from a variety of sources, including *A Dictionary of Epidemiology* (Miquel M. Porta et al. eds., 5th ed. 2008); Joseph L. Gastwirth, *Statistical Reasoning in Law and Public Policy* (1988); James K. Brewer, *Everything You Always Wanted to Know about Statistics, but Didn't Know How to Ask* (1978); and R.A. Fisher, *Statistical Methods for Research Workers* (1973).

adjustment. Methods of modifying an observed association to take into account the effect of risk factors that are not the focus of the study and that distort the observed association between the exposure being studied and the disease outcome. See also direct age adjustment, indirect age adjustment.

agent. Also, risk factor. A factor, such as a drug, microorganism, chemical substance, or form of radiation, whose presence or absence can result in the occurrence of a disease. A disease may be caused by a single agent or a number of independent alternative agents, or the combined presence of a complex of two or more factors may be necessary for the development of the disease.

alpha. The level of statistical significance chosen by a researcher to determine if any association found in a study is sufficiently unlikely to have occurred by chance (as a result of random sampling error) if the null hypothesis (no association) is true. Researchers commonly adopt an alpha of .05, but the choice is arbitrary, and other values can be justified.

alpha error. Also called Type I error and false-positive error, alpha error occurs when a researcher rejects a null hypothesis when it is actually true (i.e., when there is no association). This can occur when an apparent difference is observed between the control group and the exposed group, but the difference is not real (i.e., it occurred by chance). A common error made by lawyers, judges, and academics is to equate the level of alpha with the legal burden of proof.

association. The degree of statistical relationship between two or more events or variables. Events are said to be associated when they occur more or less frequently together than one would expect by chance. Association does not necessarily imply a causal relationship. Events are said not to have an association when the agent (or independent variable) has no apparent effect on the incidence of a disease (the dependent variable). This corresponds to a relative risk of 1.0. A negative association means that the events occur less frequently together than one would expect by chance, thereby implying a preventive or protective role for the agent (e.g., a vaccine).

attributable fraction. Also, attributable risk. The proportion of disease in exposed individuals that can be attributed to exposure to an agent, as distinguished from the proportion of disease attributed to all other causes.

attributable proportion of risk (PAR). This term has been used to denote the fraction of risk that is attributable to exposure to a substance (e.g., X percent of lung cancer is attributable to cigarettes). Synonymous terms include attributable fraction, attributable risk, etiologic fraction, population attributable risk, and risk difference. See attributable risk.

background risk of disease. Also, background rate of disease. Rate of disease in a population that has no known exposures to an alleged risk factor for the disease. For example, the background risk for all birth defects is 3–5% of live births.

beta error. Also called Type II error and false-negative error. Occurs when a researcher fails to reject a null hypothesis when it is incorrect (i.e., when there is an association). This can occur when no statistically significant difference is detected between the control group and the exposed group, but a difference does exist.

bias. Any effect at any stage of investigation or inference tending to produce results that depart systematically from the true values. In epidemiology, the term bias does not necessarily carry an imputation of prejudice or other subjective factor, such as the experimenter's desire for a particular outcome. This differs from conventional usage, in which bias refers to a partisan point of view.

biological marker. A physiological change in tissue or body fluids that occurs as a result of an exposure to an agent and that can be detected in the laboratory. Biological markers are only available for a small number of chemicals.

biological plausibility. Consideration of existing knowledge about human biology and disease pathology to provide a judgment about the plausibility that an agent causes a disease.

case-comparison study. See case-control study.

case-control study. Also, case-comparison study, case history study, case referent study, retrospective study. A study that starts with the identification of persons with a disease (or other outcome variable) and a suitable control (comparison, reference) group of persons without the disease. Such a study is often referred to as retrospective because it starts after the onset of disease and looks back to the postulated causal factors.

case group. A group of individuals who have been exposed to the disease, intervention, procedure, or other variable whose influence is being studied.

causation. As used here, an event, condition, characteristic, or agent being a necessary element of a set of other events that can produce an outcome, such as a disease. Other sets of events may also cause the disease. For example, smoking is a necessary element of a set of events that result in lung cancer, yet there are other sets of events (without smoking) that cause lung cancer. Thus, a cause may be thought of as a necessary link in at least one causal chain that

results in an outcome of interest. Epidemiologists generally speak of causation in a group context; hence, they will inquire whether an increased incidence of a disease in a cohort was “caused” by exposure to an agent.

clinical trial. An experimental study that is performed to assess the efficacy and safety of a drug or other beneficial treatment. Unlike observational studies, clinical trials can be conducted as experiments and use randomization, because the agent being studied is thought to be beneficial.

cohort. Any designated group of persons followed or traced over a period of time to examine health or mortality experience.

cohort study. The method of epidemiologic study in which groups of individuals can be identified who are, have been, or in the future may be differentially exposed to an agent or agents hypothesized to influence the incidence of occurrence of a disease or other outcome. The groups are observed to find out if the exposed group is more likely to develop disease. The alternative terms for a cohort study (concurrent study, followup study, incidence study, longitudinal study, prospective study) describe an essential feature of the method, which is observation of the population for a sufficient number of person-years to generate reliable incidence or mortality rates in the population subsets. This generally implies study of a large population, study for a prolonged period (years), or both.

confidence interval. A range of values calculated from the results of a study within which the true value is likely to fall; the width of the interval reflects random error. Thus, if a confidence level of .95 is selected for a study, 95% of similar studies would result in the true relative risk falling within the confidence interval. The width of the confidence interval provides an indication of the precision of the point estimate or relative risk found in the study; the narrower the confidence interval, the greater the confidence in the relative risk estimate found in the study. Where the confidence interval contains a relative risk of 1.0, the results of the study are not statistically significant.

confounding factor. Also, confounder. A factor that is both a risk factor for the disease and a factor associated with the exposure of interest. Confounding refers to a situation in which an association between an exposure and outcome is all or partly the result of a factor that affects the outcome but is unaffected by the exposure.

control group. A comparison group comprising individuals who have not been exposed to the disease, intervention, procedure, or other variable whose influence is being studied.

cross-sectional study. A study that examines the relationship between disease and variables of interest as they exist in a population at a given time. A cross-sectional study measures the presence or absence of disease and other variables in each member of the study population. The data are analyzed to

determine if there is a relationship between the existence of the variables and disease. Because cross-sectional studies examine only a particular moment in time, they reflect the prevalence (existence) rather than the incidence (rate) of disease and can offer only a limited view of the causal association between the variables and disease. Because exposures to toxic agents often change over time, cross-sectional studies are rarely used to assess the toxicity of exogenous agents.

data dredging. Jargon that refers to results identified by researchers who, after completing a study, pore through their data seeking to find any associations that may exist. In general, good research practice is to identify the hypotheses to be investigated in advance of the study; hence, data dredging is generally frowned on. In some cases, however, researchers conduct exploratory studies designed to generate hypotheses for further study.

demographic study. See ecological study.

dependent variable. The outcome that is being assessed in a study based on the effect of another characteristic—the independent variable. Epidemiologic studies attempt to determine whether there is an association between the independent variable (exposure) and the dependent variable (incidence of disease).

differential misclassification. A form of bias that is due to the misclassification of individuals or a variable of interest when the misclassification varies among study groups. This type of bias occurs when, for example, it is incorrectly determined that individuals in a study are unexposed to the agent being studied when in fact they are exposed. See nondifferential misclassification.

direct adjustment. A technique used to eliminate any difference between two study populations based on age, sex, or some other parameter that might result in confounding. Direct adjustment entails comparison of the study group with a large reference population to determine the expected rates based on the characteristic, such as age, for which adjustment is being performed.

dose. Generally refers to the intensity or magnitude of exposure to an agent multiplied by the duration of exposure. Dose may be used to refer only to the intensity of exposure.

dose–response relationship. A relationship in which a change in amount, intensity, or duration of exposure to an agent is associated with a change—either an increase or a decrease—in risk of disease.

double blinding. A method used in experimental studies in which neither the individuals being studied nor the researchers know during the study whether any individual has been assigned to the exposed or control group. Double blinding is designed to prevent knowledge of the group to which the individual was assigned from biasing the outcome of the study.

ecological fallacy. Also, aggregation bias, ecological bias. An error that occurs from inferring that a relationship that exists for groups is also true for individuals. For example, if a country with a higher proportion of fishermen also has a higher rate of suicides, then inferring that fishermen must be more likely to commit suicide is an ecological fallacy.

ecological study. Also, demographic study. A study of the occurrence of disease based on data from populations, rather than from individuals. An ecological study searches for associations between the incidence of disease and suspected disease-causing agents in the studied populations. Researchers often conduct ecological studies by examining easily available health statistics, making these studies relatively inexpensive in comparison with studies that measure disease and exposure to agents on an individual basis.

epidemiology. The study of the distribution and determinants of disease or other health-related states and events in populations and the application of this study to control of health problems.

error. Random error (sampling error) is the error that is due to chance when the result obtained for a sample differs from the result that would be obtained if the entire population (universe) were studied.

etiologic factor. An agent that plays a role in causing a disease.

etiology. The cause of disease or other outcome of interest.

experimental study. A study in which the researcher directly controls the conditions. Experimental epidemiology studies (also clinical studies) entail random assignment of participants to the exposed and control groups (or some other method of assignment designed to minimize differences between the groups).

exposed, exposure. In epidemiology, the exposed group (or the exposed) is used to describe a group whose members have been exposed to an agent that may be a cause of a disease or health effect of interest, or possess a characteristic that is a determinant of a health outcome.

false-negative error. See beta error.

false-positive error. See alpha error.

followup study. See cohort study.

general causation. Issue of whether an agent increases the incidence of disease in a group and not whether the agent caused any given individual's disease. Because of individual variation, a toxic agent generally will not cause disease in every exposed individual.

generalizable. When the results of a study are applicable to populations other than the study population, such as the general population.

in vitro. Within an artificial environment, such as a test tube (e.g., the cultivation of tissue in vitro).

in vivo. Within a living organism (e.g., the cultivation of tissue in vivo).

incidence rate. The number of people in a specified population falling ill from a particular disease during a given period. More generally, the number of new events (e.g., new cases of a disease in a defined population) within a specified period of time.

incidence study. See cohort study.

independent variable. A characteristic that is measured in a study and that is suspected to have an effect on the outcome of interest (the dependent variable). Thus, exposure to an agent is measured in a cohort study to determine whether that independent variable has an effect on the incidence of disease, which is the dependent variable.

indirect adjustment. A technique employed to minimize error that might result when comparing two populations because of differences in age, sex, or another parameter that may independently affect the rate of disease in the populations. The incidence of disease in a large reference population, such as all residents of a country, is calculated for each subpopulation (based on the relevant parameter, such as age). Those incidence rates are then applied to the study population with its distribution of persons to determine the overall incidence rate for the study population, which provides a standardized mortality or morbidity ratio (often referred to as SMR).

inference. The intellectual process of making generalizations from observations. In statistics, the development of generalizations from sample data, usually with calculated degrees of uncertainty.

information bias. Also, observational bias. Systematic error in measuring data that results in differential accuracy of information (such as exposure status) for comparison groups.

interaction. When the magnitude or direction (positive or negative) of the effect of one risk factor differs depending on the presence or level of the other. In interaction, the effect of two risk factors together is different (greater or less) than the sum of their individual effects.

meta-analysis. A technique used to combine the results of several studies to enhance the precision of the estimate of the effect size and reduce the plausibility that the association found is due to random sampling error. Meta-analysis is best suited to pooling results from randomly controlled experimental studies, but if carefully performed, it also may be useful for observational studies.

misclassification bias. The erroneous classification of an individual in a study as exposed to the agent when the individual was not, or incorrectly classifying a study individual with regard to disease. Misclassification bias may exist in all study groups (nondifferential misclassification) or may vary among groups (differential misclassification).

morbidity rate. State of illness or disease. Morbidity rate may refer to either the incidence rate or prevalence rate of disease.

mortality rate. Proportion of a population that dies of a disease or of all causes. The numerator is the number of individuals dying; the denominator is the total population in which the deaths occurred. The unit of time is usually a calendar year.

model. A representation or simulation of an actual situation. This may be either (1) a mathematical representation of characteristics of a situation that can be manipulated to examine consequences of various actions; (2) a representation of a country's situation through an "average region" with characteristics resembling those of the whole country; or (3) the use of animals as a substitute for humans in an experimental system to ascertain an outcome of interest.

multivariate analysis. A set of techniques used when the variation in several variables has to be studied simultaneously. In statistics, any analytical method that allows the simultaneous study of two or more independent factors or variables.

nondifferential misclassification. Error due to misclassification of individuals or a variable of interest into the wrong category when the misclassification varies among study groups. The error may result from limitations in data collection, may result in bias, and will often produce an underestimate of the true association. See differential misclassification.

null hypothesis. A hypothesis that states that there is no true association between a variable and an outcome. At the outset of any observational or experimental study, the researcher must state a proposition that will be tested in the study. In epidemiology, this proposition typically addresses the existence of an association between an agent and a disease. Most often, the null hypothesis is a statement that exposure to Agent A does not increase the occurrence of Disease D. The results of the study may justify a conclusion that the null hypothesis (no association) has been disproved (e.g., a study that finds a strong association between smoking and lung cancer). A study may fail to disprove the null hypothesis, but that alone does not justify a conclusion that the null hypothesis has been proved.

observational study. An epidemiologic study in situations in which nature is allowed to take its course, without intervention from the investigator. For example, in an observational study the subjects of the study are permitted to determine their level of exposure to an agent.

odds ratio (OR). Also, cross-product ratio, relative odds. The ratio of the odds that a case (one with the disease) was exposed to the odds that a control (one without the disease) was exposed. For most purposes the odds ratio from a case-control study is quite similar to a risk ratio from a cohort study.

***p* (probability), *p*-value.** The *p*-value is the probability of getting a value of the test outcome equal to or more extreme than the result observed, given that the null hypothesis is true. The letter *p*, followed by the abbreviation “n.s.” (not significant) means that $p > .05$ and that the association was not statistically significant at the .05 level of significance. The statement “ $p < .05$ ” means that *p* is less than 5%, and, by convention, the result is deemed statistically significant. Other significance levels can be adopted, such as .01 or .1. The lower the *p*-value, the less likely that random error would have produced the observed relative risk if the true relative risk is 1.

pathognomonic. When an agent must be present for a disease to occur. Thus, asbestos is a pathognomonic agent for asbestosis. See signature disease.

placebo controlled. In an experimental study, providing an inert substance to the control group, so as to keep the control and exposed groups ignorant of their status.

power. The probability that a difference of a specified amount will be detected by the statistical hypothesis test, given that a difference exists. In less formal terms, power is like the strength of a magnifying lens in its capability to identify an association that truly exists. Power is equivalent to one minus Type II error. This is sometimes stated as $\text{Power} = 1 - \beta$.

prevalence. The percentage of persons with a disease in a population at a specific point in time.

prospective study. A study in which two groups of individuals are identified: (1) individuals who have been exposed to a risk factor and (2) individuals who have not been exposed. Both groups are followed for a specified length of time, and the proportion that develops disease in the first group is compared with the proportion that develops disease in the second group. See cohort study.

random. The term implies that an event is governed by chance. See randomization.

randomization. Assignment of individuals to groups (e.g., for experimental and control regimens) by chance. Within the limits of chance variation, randomization should make the control group and experimental group similar at the start of an investigation and ensure that personal judgment and prejudices of the investigator do not influence assignment. Randomization should not be confused with haphazard assignment. Random assignment follows a predetermined plan that usually is devised with the aid of a table of random numbers. Randomization cannot ethically be used where the exposure is known to cause harm (e.g., cigarette smoking).

randomized trial. See clinical trial.

recall bias. Systematic error resulting from differences between two groups in a study in accuracy of memory. For example, subjects who have a disease may recall exposure to an agent more frequently than subjects who do not have the disease.

relative risk (RR). The ratio of the risk of disease or death among people exposed to an agent to the risk among the unexposed. For instance, if 10% of all people exposed to a chemical develop a disease, compared with 5% of people who are not exposed, the disease occurs twice as frequently among the exposed people. The relative risk is $10\%/5\% = 2$. A relative risk of 1 indicates no association between exposure and disease.

research design. The procedures and methods, predetermined by an investigator, to be adhered to in conducting a research project.

risk. A probability that an event will occur (e.g., that an individual will become ill or die within a stated period of time or by a certain age).

risk difference (RD). The difference between the proportion of disease in the exposed population and the proportion of disease in the unexposed population. $-1.0 \leq RD \leq 1.0$.

sample. A selected subset of a population. A sample may be random or nonrandom.

sample size. The number of subjects who participate in a study.

secular-trend study. Also, time-line study. A study that examines changes over a period of time, generally years or decades. Examples include the decline of tuberculosis mortality and the rise, followed by a decline, in coronary heart disease mortality in the United States in the past 50 years.

selection bias. Systematic error that results from individuals being selected for the different groups in an observational study who have differences other than the ones that are being examined in the study.

sensitivity. Measure of the accuracy of a diagnostic or screening test or device in identifying disease (or some other outcome) when it truly exists. For example, assume that we know that 20 women in a group of 1000 women have cervical cancer. If the entire group of 1000 women is tested for cervical cancer and the screening test only identifies 15 (of the known 20) cases of cervical cancer, the screening test has a sensitivity of $15/20$, or 75%. Also see specificity.

signature disease. A disease that is associated uniquely with exposure to an agent (e.g., asbestosis and exposure to asbestos). See also pathognomonic.

significance level. A somewhat arbitrary level selected to minimize the risk that an erroneous positive study outcome that is due to random error will be accepted as a true association. The lower the significance level selected, the less likely that false-positive error will occur.

specific causation. Whether exposure to an agent was responsible for a given individual's disease.

specificity. Measure of the accuracy of a diagnostic or screening test in identifying those who are disease-free. Once again, assume that 980 women out of a group of 1000 women do not have cervical cancer. If the entire group of 1000 women is screened for cervical cancer and the screening test only iden-

tifies 900 women without cervical cancer, the screening test has a specificity of 900/980, or 92%.

standardized morbidity ratio (SMR). The ratio of the incidence of disease observed in the study population to the incidence of disease that would be expected if the study population had the same incidence of disease as some selected reference population.

standardized mortality ratio (SMR). The ratio of the incidence of death observed in the study population to the incidence of death that would be expected if the study population had the same incidence of death as some selected standard or known population.

statistical significance. A term used to describe a study result or difference that exceeds the Type I error rate (or p -value) that was selected by the researcher at the outset of the study. In formal significance testing, a statistically significant result is unlikely to be the result of random sampling error and justifies rejection of the null hypothesis. Some epidemiologists believe that formal significance testing is inferior to using a confidence interval to express the results of a study. Statistical significance, which addresses the role of random sampling error in producing the results found in the study, should not be confused with the importance (for public health or public policy) of a research finding.

stratification. Separating a group into subgroups based on specified criteria, such as age, gender, or socioeconomic status. Stratification is used both to control for the possibility of confounding (by separating the studied populations based on the suspected confounding factor) and when there are other known factors that affect the disease under study. Thus, the incidence of death increases with age, and a study of mortality might use stratification of the cohort and control groups based on age.

study design. See research design.

systematic error. See bias.

teratogen. An agent that produces abnormalities in the embryo or fetus by disturbing maternal health or by acting directly on the fetus in utero.

teratogenicity. The capacity for an agent to produce abnormalities in the embryo or fetus.

threshold phenomenon. A certain level of exposure to an agent below which disease does not occur and above which disease does occur.

time-line study. See secular-trend study.

toxicology. The science of the nature and effects of poisons. Toxicologists study adverse health effects of agents on biological organisms, such as live animals and cells. Studies of humans are performed by epidemiologists.

toxic substance. A substance that is poisonous.

true association. Also, real association. The association that really exists between exposure to an agent and a disease and that might be found by a perfect (but nonetheless nonexistent) study.

Type I error. Rejecting the null hypothesis when it is true. See alpha error.

Type II error. Failing to reject the null hypothesis when it is false. See beta error.

validity. The degree to which a measurement measures what it purports to measure; the accuracy of a measurement.

variable. Any attribute, condition, or other characteristic of subjects in a study that can have different numerical characteristics. In a study of the causes of heart disease, blood pressure and dietary fat intake are variables that might be measured.

References on Epidemiology

- Causal Inferences (Kenneth J. Rothman ed., 1988).
- William G. Cochran, *Sampling Techniques* (1977).
- A Dictionary of Epidemiology (John M. Last et al. eds., 5th ed. 2008).
- Anders Ahlbom & Steffan Norell, *Introduction to Modern Epidemiology* (2d ed. 1990).
- Robert C. Elston & William D. Johnson, *Basic Biostatistics for Geneticists and Epidemiologists* (2008)
- Encyclopedia of Epidemiology (Sarah E. Boslaugh ed., 2008).
- Joseph L. Fleiss et al., *Statistical Methods for Rates and Proportions* (3d ed. 2003).
- Leon Gordis, *Epidemiology* (4th ed. 2009).
- Morton Hunt, *How Science Takes Stock: The Story of Meta-Analysis* (1997).
- International Agency for Research on Cancer (IARC), *Interpretation of Negative Epidemiologic Evidence for Carcinogenicity* (N.J. Wald & R. Doll eds., 1985).
- Harold A. Kahn & Christopher T. Sempos, *Statistical Methods in Epidemiology* (1989).
- David E. Lilienfeld, *Overview of Epidemiology*, 3 *Shepard's Expert & Sci. Evid. Q.* 25 (1995).
- David E. Lilienfeld & Paul D. Stolley, *Foundations of Epidemiology* (3d ed. 1994).
- Marcello Pagano & Kimberlee Gauvreau, *Principles of Biostatistics* (2d ed. 2000).
- Pharmacoepidemiology (Brian L. Strom ed., 4th ed. 2005).
- Richard K. Riegelman & Robert A. Hirsch, *Studying a Study and Testing a Test: How to Read the Health Science Literature* (5th ed. 2005).
- Bernard Rosner, *Fundamentals of Biostatistics* (6th ed. 2006).
- Kenneth J. Rothman et al., *Modern Epidemiology* (3d ed. 2008).
- David A. Savitz, *Interpreting Epidemiologic Evidence: Strategies for Study Design and Analysis* (2003).
- James J. Schlesselman, *Case-Control Studies: Design, Conduct, Analysis* (1982).
- Lisa M. Sullivan, *Essentials of Biostatistics* (2008).
- Mervyn Susser, *Epidemiology, Health and Society: Selected Papers* (1987).

References on Law and Epidemiology

- American Law Institute, *Reporters' Study on Enterprise Responsibility for Personal Injury* (1991).
- Bert Black & David H. Hollander, Jr., *Unraveling Causation: Back to the Basics*, 3 *U. Balt. J. Envtl. L.* 1 (1993).
- Bert Black & David Lilienfeld, *Epidemiologic Proof in Toxic Tort Litigation*, 52 *Fordham L. Rev.* 732 (1984).

- Gerald Boston, *A Mass-Exposure Model of Toxic Causation: The Content of Scientific Proof and the Regulatory Experience*, 18 Colum. J. Envtl. L. 181 (1993).
- Vincent M. Brannigan et al., *Risk, Statistical Inference, and the Law of Evidence: The Use of Epidemiological Data in Toxic Tort Cases*, 12 Risk Analysis 343 (1992).
- Troyen Brennan, *Causal Chains and Statistical Links: The Role of Scientific Uncertainty in Hazardous-Substance Litigation*, 73 Cornell L. Rev. 469 (1988).
- Troyen Brennan, *Helping Courts with Toxic Torts: Some Proposals Regarding Alternative Methods for Presenting and Assessing Scientific Evidence in Common Law Courts*, 51 U. Pitt. L. Rev. 1 (1989).
- Philip Cole, *Causality in Epidemiology, Health Policy, and Law*, 27 Envtl. L. Rep. 10,279 (June 1997).
- Comment, *Epidemiologic Proof of Probability: Implementing the Proportional Recovery Approach in Toxic Exposure Torts*, 89 Dick. L. Rev. 233 (1984).
- George W. Conk, *Against the Odds: Proving Causation of Disease with Epidemiological Evidence*, 3 Shepard's Expert & Sci. Evid. Q. 85 (1995).
- Carl F. Cranor, *Toxic Torts: Science, Law, and the Possibility of Justice* (2006).
- Carl F. Cranor et al., *Judicial Boundary Drawing and the Need for Context-Sensitive Science in Toxic Torts After Daubert v. Merrell Dow Pharmaceuticals, Inc.*, 16 Va. Envtl. L.J. 1 (1996).
- Richard Delgado, *Beyond Sindell: Relaxation of Cause-in-Fact Rules for Indeterminate Plaintiffs*, 70 Cal. L. Rev. 881 (1982).
- Michael Dore, *A Commentary on the Use of Epidemiological Evidence in Demonstrating Cause-in-Fact*, 7 Harv. Envtl. L. Rev. 429 (1983).
- Jean Macchiaroli Eggen, *Toxic Torts, Causation, and Scientific Evidence After Daubert*, 55 U. Pitt. L. Rev. 889 (1994).
- Daniel A. Farber, *Toxic Causation*, 71 Minn. L. Rev. 1219 (1987).
- Heidi Li Feldman, *Science and Uncertainty in Mass Exposure Litigation*, 74 Tex. L. Rev. 1 (1995).
- Stephen E. Fienberg et al., *Understanding and Evaluating Statistical Evidence in Litigation*, 36 Jurimetrics J. 1 (1995).
- Joseph L. Gastwirth, *Statistical Reasoning in Law and Public Policy* (1988).
- Herman J. Gibb, *Epidemiology and Cancer Risk Assessment*, in *Fundamentals of Risk Analysis and Risk Management* 23 (Vlasta Molak ed., 1997).
- Steve Gold, Note, *Causation in Toxic Torts: Burdens of Proof, Standards of Persuasion and Statistical Evidence*, 96 Yale L.J. 376 (1986).
- Leon Gordis, *Epidemiologic Approaches for Studying Human Disease in Relation to Hazardous Waste Disposal Sites*, 25 Hous. L. Rev. 837 (1988).
- Michael D. Green, *Expert Witnesses and Sufficiency of Evidence in Toxic Substances Litigation: The Legacy of Agent Orange and Bendectin Litigation*, 86 Nw. U. L. Rev. 643 (1992).
- Michael D. Green, *The Future of Proportional Liability*, in *Exploring Tort Law* (Stuart Madden ed., 2005).

- Sander Greenland, *The Need for Critical Appraisal of Expert Witnesses in Epidemiology and Statistics*, 39 Wake Forest L. Rev. 291 (2004).
- Khristine L. Hall & Ellen Silbergeld, *Reappraising Epidemiology: A Response to Mr. Dore*, 7 Harv. Envtl. L. Rev. 441 (1983).
- Jay P. Kesan, *Drug Development: Who Knows Where the Time Goes?: A Critical Examination of the Post-Daubert Scientific Evidence Landscape*, 52 Food Drug Cosm. L.J. 225 (1997).
- Jay P. Kesan, *An Autopsy of Scientific Evidence in a Post-Daubert World*, 84 Geo. L. Rev. 1985 (1996).
- Constantine Kokkoris, Comment, *DeLuca v. Merrell Dow Pharmaceuticals, Inc.: Statistical Significance and the Novel Scientific Technique*, 58 Brook. L. Rev. 219 (1992).
- James P. Leape, *Quantitative Risk Assessment in Regulation of Environmental Carcinogens*, 4 Harv. Envtl. L. Rev. 86 (1980).
- David E. Lilienfeld, *Overview of Epidemiology*, 3 Shepard's Expert & Sci. Evid. Q. 23 (1995).
- Junius McElveen, Jr., & Pamela Eddy, *Cancer and Toxic Substances: The Problem of Causation and the Use of Epidemiology*, 33 Clev. St. L. Rev. 29 (1984).
- Modern Scientific Evidence: The Law and Science of Expert Testimony (David L. Faigman et al. eds., 2009–2010).
- Note, *Development in the Law—Confronting the New Challenges of Scientific Evidence*, 108 Harv. L. Rev. 1481 (1995).
- Susan R. Poulter, *Science and Toxic Torts: Is There a Rational Solution to the Problem of Causation?* 7 High Tech. L.J. 189 (1992).
- Jon Todd Powell, Comment, *How to Tell the Truth with Statistics: A New Statistical Approach to Analyzing the Data in the Aftermath of Daubert v. Merrell Dow Pharmaceuticals*, 31 Hous. L. Rev. 1241 (1994).
- Restatement (Third) of Torts: Liability for Physical and Emotional Harm § 28, cmt. c & rptrs. note (2010).
- David Rosenberg, *The Causal Connection in Mass Exposure Cases: A Public Law Vision of the Tort System*, 97 Harv. L. Rev. 849 (1984).
- Joseph Sanders, *The Bendectin Litigation: A Case Study in the Life-Cycle of Mass Torts*, 43 Hastings L.J. 301 (1992).
- Joseph Sanders, *Scientific Validity, Admissibility, and Mass Torts After Daubert*, 78 Minn. L. Rev. 1387 (1994).
- Joseph Sanders & Julie Machal-Fulks, *The Admissibility of Differential Diagnosis to Prove Causation in Toxic Tort Cases: The Interplay of Adjective and Substantive Law*, 64 L. & Contemp. Probs. 107 (2001).
- Palma J. Strand, *The Inapplicability of Traditional Tort Analysis to Environmental Risks: The Example of Toxic Waste Pollution Victim Compensation*, 35 Stan. L. Rev. 575 (1983).
- Richard W. Wright, *Causation in Tort Law*, 73 Cal. L. Rev. 1735 (1985).

Appendix C (2 pages)

GRADE working group

[Home](#) [Introduction](#) [Toolbox](#) [Publications](#) [Member login](#) [Links](#) [Contact](#)

Introduction

small logo

[Learn more](#) [FAQ](#) [Organizations](#) [Downloads](#) [Courses](#) [About us](#)

Grading the quality of evidence and the strength of recommendations

Judgments about evidence and recommendations in healthcare are complex. For example, those making recommendations must decide between recommending selective serotonin reuptake inhibitors (SSRI's) and tricyclics for the treatment of moderate depression must agree on which outcomes to consider, which evidence to include for each outcome, how to assess the quality of that evidence, and how to determine if SSRI's do more good than harm compared with tricyclics. Because resources are always limited and money that is allocated to treating depression cannot be spent on other worthwhile interventions, they may also need to decide whether any incremental health benefits are worth the additional costs.

Systematic reviews of the effects of healthcare provide essential, but not sufficient information for making well informed decisions. Reviewers and people who use reviews draw conclusions about the quality of the evidence, either implicitly or explicitly. Such judgments guide subsequent decisions. For example, clinical actions are likely to differ depending on whether one concludes that the evidence that warfarin reduces the risk of stroke in patients with atrial fibrillation is convincing (high quality) or that it is unconvincing (low quality).

Similarly, practice guidelines and people who use them draw conclusions about the strength of recommendations, either implicitly or explicitly. Using the same example, a guideline that recommends that patients with atrial fibrillation should be treated may suggest that all patients definitely should be treated or that patients should probably be treated, implying that treatment may not be warranted in all patients.

A systematic and explicit approach to making judgments such as these can help to prevent errors, facilitate critical appraisal of these judgments, and can help to improve communication of this information. Since the 1970's a growing number of organizations have employed various systems to grade the quality (level) of evidence and the strength of recommendations. Unfortunately, different organizations use different systems to grade evidence and recommendations. The same evidence and recommendation could be graded as "II-2, B", "C+, 1", or "strong evidence, strongly recommended" depending on which system is used. This is confusing and impedes effective communication.

To learn more about the GRADE approach, please see our [FAQ](#) or our [publication section](#) and the following Criteria for applying and using GRADE:

Criteria for applying or using GRADE [\[pdf\]](#)

One of the aims of the GRADE Working Group is to reduce unnecessary confusion arising from multiple systems for grading evidence and recommendations. To avoid adding to this confusion by having multiple variations of the GRADE system we suggest that the criteria below should be met when saying that the GRADE system was used. Also, while users may believe there may be good reasons for modifying the GRADE system, we discourage the use

of “modified GRADE approaches” that differ substantially from the approach described by the GRADE Working Group.

On the other hand, we encourage and welcome constructive criticism of the GRADE approach, suggestions for improvements, and involvement in the GRADE Working Group. As most scientific approaches to advancing healthcare, the GRADE approach will continue to evolve in response to new evidence and to meet the needs of systematic review authors, guideline developers and other users.

Suggested criteria for stating that the GRADE system was used:

1. “Quality of evidence” should be defined consistently with one of the two definitions (for guidelines or for systematic reviews) used by the GRADE Working Group.
2. Explicit consideration should be given to each of the GRADE criteria for assessing the quality of evidence (risk of bias/study limitations, directness, consistency of results, precision, publication bias, magnitude of the effect, dose-response gradient, influence of residual plausible confounding and bias “antagonistic bias”) although different terminology may be used.
3. The overall quality of evidence should be assessed for each important outcome and expressed using four (e.g. high, moderate, low, very low) or, if justified, three (e.g. high, moderate, and very low and low combined into low) categories based on definitions for each category that are consistent with the definitions used by the GRADE Working Group.
4. Evidence summaries (narrative or in table format) should be used as the basis for judgements about the quality of evidence and the strength of recommendations. Ideally, full evidence profiles suggested by the GRADE Working Group should be used and these should be based on systematic reviews. At a minimum, the evidence that was assessed and the methods that were used to identify and appraise that evidence should be clearly described. In particular, reasons for up and downgrading should be described transparently.
5. Explicit consideration should be given to each of the GRADE criteria for assessing the strength of a recommendation (the balance of desirable and undesirable consequences, quality of evidence, values and preferences, and resource use) and a general approach should be reported (e.g. if and how costs were considered, whose values and preferences were assumed, etc.).
6. The strength of recommendations should be expressed using two categories (weak/conditional and strong) for or against a management option and the definitions for each category should be consistent with those used by the GRADE Working Group. Different terminology to express weak/conditional and strong recommendations may be used, although the interpretation and implications should be preserved.
7. Decisions about the strength of the recommendations should ideally be transparently reported.

Wed, 8/1/2012, 2:00 PM - 3:50 PM

<http://www.amstat.org/meetings/jsm/2012/onlineprogram/ActivityDetails.cfm?SessionID=207510>

Are Fine Particulates Killing Californians? — Invited Papers

Section on Risk Analysis , Section on Survey Research Methods , Section on Statistics and the Environment , Section for Statistical Programmers and Analysts , Section on Statistics in Epidemiology

Organizer(s): Michael E Ginevan, M.E. Ginevan & Associates

Chair(s): Michael E Ginevan, M.E. Ginevan & Associates

2:05 PM [Particulate Matter is Not Killing Californians](#) — **James E. Enstrom, University of California at Los Angeles**

2:25 PM [A Closer Look at Air Pollution-Mortality Relationships for California Members of the American Cancer Society Cohort](#) — **Frederick W. Lipfert, Environmental Consultant** ; S. Stanley Young, National Institute of Statistical Sciences

2:45 PM [Assessing Variable Importance in an Environmental Observational Study](#) — **S. Stanley Young, National Institute of Statistical Sciences** ; Jesse Q. Xia, National Institute of Statistical Sciences

3:05 PM [Improving the Scientific Advice Provided by the Clean Air Scientific Advisory PM Subcommittee](#) — **Robert F. Phalen, University of California at Irvine**

3:25 PM **Discussant:** Michael E Ginevan, M.E. Ginevan & Associates

3:45 PM **Floor Discussion**

01 Particulate Matter is Not Killing Californians

Author(s): James E. Enstrom*+

Companies: University of California at Los Angeles

Address: BOX 951772, A1-295 CHS, Los Angeles, CA, 90095-1772,

Keywords: [epidemiology](#) ; [particulate matter](#) ; [mortality](#) ; [causality](#) ; [statistics](#) ; [California](#)

Abstract: There is now overwhelming epidemiologic evidence that particulate matter (PM), both fine particulate matter (PM_{2.5}) and coarse particulate matter (PM₁₀), is not related to total mortality in California. I will examine all the long-term PM epidemiologic cohort studies in California, and discuss the ways the findings from these studies have been used and/or ignored. I will discuss the limitations of these studies: lack of access to key databases; the ecological fallacy; failure to consider other pollutants; failure to satisfy causality criteria; and failure to consider other competing health risks. Also, ethical issues underlying much of PM_{2.5} epidemiology will be discussed. I will make a strong case that PM_{2.5} is not killing Californians and that there is not a scientific or public health basis for the many of the existing and proposed regulations designed to reduce PM levels in California. Finally, I will make the case that PM health effects and regulations must be put into perspective with other factors that influence health in California, given the low age-adjusted total death rate in this state.

Particulate Matter is Not Killing Californians

James E. Enstrom, Ph.D., M.P.H.
UCLA School of Public Health
Los Angeles, CA 90095-1772

[http://www.scientificintegrityinstitute.org/
jenstrom@ucla.edu](http://www.scientificintegrityinstitute.org/jenstrom@ucla.edu)

September 28, 2012

Abstract

There is now overwhelming epidemiologic evidence that particulate matter (PM), both fine particulate matter (PM_{2.5}) and course particulate matter (PM₁₀), is not related to total mortality in California. I will examine all the long-term PM epidemiologic cohort studies in California, and discuss the ways the findings from these studies have been used and/or ignored. I will discuss the limitations of these studies: lack of access to key databases; the ecological fallacy; failure to consider other pollutants; failure to satisfy causality criteria; and failure to consider other competing health risks. Also, ethical issues underlying much of PM_{2.5} epidemiology will be discussed. I will make a strong case that PM_{2.5} is not killing Californians and that there is not a scientific or public health basis for the many of the existing and proposed regulations designed to reduce PM levels in California. Finally, I will make the case that PM health effects and regulations must be put into perspective with other factors that influence health in California, given the low age-adjusted total death rate in this state.

Key Words: epidemiology, particulate matter, mortality, causality, statistics, California

1. Background

1.1 Relationship of PM_{2.5} Epidemiology to EPA, CARB, and AQMD

This paper focuses on particulate matter (PM) epidemiology in California. PM consists of fine particulates (PM_{2.5}), defined to have particle size <2.5 μm in diameter, and course particulates (PM₁₀), defined to have a particle size <10 μm in diameter. PM_{2.5} is generated mainly by combustion processes, such as, forest fires, agricultural dust, industrial combustion, and diesel engines. PM_{2.5} epidemiology played a major role in the US Environmental Protection Agency (EPA) establishment of the 1997 National Ambient Air Quality Standard (NAAQS) for PM_{2.5} (<http://www.epa.gov/air/criteria.html>). EPA has recently proposed to lower the annual NAAQS for PM_{2.5} from the current level of 15 μg/m³ to 12-13 μg/m³

(<http://www.epa.gov/pm/actions.html>). The PM2.5 regulations established since 1997 have had multi-billion dollar economic impacts in the United States and California and have been highly contested (<http://science.house.gov/press-release/harris-and-brown-question-administration%E2%80%99s-environmental-cost-benefit-analyses>).

PM2.5 epidemiology has also been used by the California Air Resources Board (CARB) to establish the draconian Truck and Bus Regulation to reduce PM emissions from diesel vehicles in California (<http://www.arb.ca.gov/msprog/onrdiesel/onrdiesel.htm>). During the past five years, I have challenged the scientific and public health justifications for these regulations ([http://www.arb.ca.gov/lists/gmbond2011/2-enstrom letter to coal conerez re suspend carb diesel regs 121311.pdf](http://www.arb.ca.gov/lists/gmbond2011/2-enstrom%20letter%20to%20coal%20concern%20re%20suspend%20carb%20diesel%20regs%20121311.pdf)).

PM2.5 epidemiology is also being used by the Southern California Air Quality Management District (AQMD) in the development of the 2012 Air Quality Management Plan (AQMP) (<http://www.aqmd.gov/aqmp/2012aqmp/index.htm>). The AQMP proposes aggressive and costly emission control measures in order to reduce existing PM and ozone levels in the South Coast Air Basin (SCAB). This air basin includes about 17 million residents in Orange County and the urban portions of Los Angeles, Riverside, and San Bernardino Counties. The primary goal of the AQMP is to bring the SCAB into compliance with the NAAQS for criteria pollutants, primarily, PM2.5 and ozone.

An elevated relative risk ($RR > 1.00$) in an epidemiologic cohort study, i.e., increase in total (all cause) mortality risk for a $10 \mu\text{g}/\text{m}^3$ increase in PM2.5 level, is interpreted by EPA, CARB, and AQMD as evidence that PM2.5 “causes” “premature deaths.” Because EPA assigns a lifetime monetary value of about \$7-9 million to each “premature death,” the health benefits of preventing these deaths exceed the compliance costs of the regulations that are designed to reduce PM2.5 levels and PM2.5-related “premature deaths.” Without PM2.5-related “premature deaths” the PM2.5 regulations are not justified on a cost-benefit basis.

During the past two decades there has been extensive criticism of PM2.5 epidemiology and its use for regulation of PM by EPA, CARB, and AQMD. Five major reasons for doubting a “causal” relationship between PM2.5 and “premature deaths” are: 1) the relative risk of death due to PM2.5 is small ($RR \sim 1.10$), varies by time and place, and shows no consistent dose-response relationship; 2) confounding variables, including other pollutants, often reduce the PM2.5 effect to zero ($RR \sim 1.00$); 3) the ecological fallacy applies to all PM2.5 epidemiology because PM2.5 measurements made at selected monitoring stations are imputed to individuals living near these stations; 4) the chemical composition of PM2.5 varies greatly across the US; and 5) the major PM2.5 epidemiologic findings that have been used to establish regulations are based on secret data maintained by the American Cancer Society and Harvard University (Krewski 2000), that is not accessible for independent reanalysis.

1.2 Major Lectures on PM2.5 and Mortality in California by Enstrom

The above epidemiologic issues are too complex to fully address in this paper. Additional relevant information can be found in the following major lectures that I have given since 2010, often in conjunction with other experts on this subject:

February 26, 2010 CARB Symposium “Estimating Premature Deaths from Long-term Exposure to PM_{2.5}, with Enstrom talk “Critique of CARB Diesel Science, 1998-2010” (http://www.arb.ca.gov/research/health/pm-mort/pm-mort-ws_02-26-10.htm) (<http://www.arb.ca.gov/research/health/pm-mort/enstrom.pdf>)

November 28, 2011 UCLA Institute of the Environment Enstrom Seminar "Does Fine Particulate Matter Kill Californians? An Epidemiologic and Regulatory Controversy" (<http://www.environment.ucla.edu/calendar/showevent.asp?eventid=667>) and ([http://www.arb.ca.gov/lists/gmbond2011/3-
ioes_seminar_does_particulate_matter_kill_californians_enstrom_112811.pdf](http://www.arb.ca.gov/lists/gmbond2011/3-ioes_seminar_does_particulate_matter_kill_californians_enstrom_112811.pdf))

April 24, 2012 Dose-Response 2012 Conference Enstrom Lecture "Pseudoscientific Aspects of Fine Particulate Matter Epidemiology, 1993-2012" (http://dose-response.org/conference/2012/pdf/Enstrom_Dose_Response_Fine_Part particulate.pdf)

August 1, 2012 American Statistical Association Joint Statistical Meeting Session "Are Fine Particulates Killing Californians?" with title talk by Enstrom (<http://www.amstat.org/meetings/jsm/2012/onlineprogram/ActivityDetails.cfm?SessionID=207510>) and (<http://www.scientificintegrityinstitute.org/ASA080112.pdf>)

2. PM_{2.5} and Total Mortality in California

2.1 California-specific Epidemiologic Results Summarized

Table 1 summarizes ten separate analyses of five major California cohorts that have found no relationship between PM_{2.5} and total mortality. References to these analyses are cited in the table and listed at the end of this paper and additional details are provided at this link (<http://www.scientificintegrityinstitute.org/Enstrom081512.pdf>). Included in Table 1 is an analysis limited to the Los Angeles area (Jerrett 2005). Table 2 summarizes five separate analyses of three of the major California cohorts. These analyses have found no relationship between PM₁₀ and total mortality. There are no statewide cohort analyses that show a positive relationship between PM (PM_{2.5} and PM₁₀) and total mortality in California. Indeed, three of these analyses (Jerrett 2011, Lipsett 2011, Ostro 2011), funded by CARB and AQMD, found no relationship between any criteria pollutant and total mortality in California.

The first published evidence of no PM_{2.5} mortality risk in California is contained in the July 2000 Health Effects Institute (HEI) Reanalysis Report (Krewski 2000). Figure 21, a U.S. map of “Fine Particulates and Mortality Risk,” indicates no excess mortality risk in California. Figure 5 provides further evidence of the geographic variation in PM_{2.5} mortality risk, with Fresno (city #3) ranking second lowest in risk among 49 cities and Los Angeles (city #39) ranking fifth lowest in risk (<http://www.scientificintegrityinstitute.org/HEIFigure5093010.pdf>). Figure 1 below reproduces Figure 21 and Figure 5 with a city number assigned to each data point. The null California PM_{2.5} mortality risk findings in Figure 21 were confirmed in the August 31, 2010 letter from Krewski to HEI (Krewski 2010).

2.2 Misrepresentation of PM_{2.5} and Mortality in California by CARB

My December 15, 2005 *Inhalation Toxicology* paper, “Fine Particulate Air Pollution and Total Mortality Among Elderly Californians, 1973–2002” (Enstrom 2005), found no relationship between PM_{2.5} and mortality in California during 1983–2002. This is the first, largest, and most detailed peer reviewed journal publication that focuses on the relationship between PM_{2.5} and total mortality in California. Enstrom 2005 appeared just after the November 2005 *Epidemiology* paper “Spatial Analysis of Air Pollution and Mortality in Los Angeles” (Jerrett 2005), which found an unusually large relative risk between PM_{2.5} and mortality in the Los Angeles basin during 1982–2000. The finding is in direct contrast to the low absolute PM_{2.5} mortality risk for Los Angeles found in Figure 21. These conflicting findings need to be resolved with further analysis.

Enstrom 2005 was submitted to CARB health effects scientist Linda Smith on January 9, 2006 (http://www.arb.ca.gov/planning/gmerp/dec1plan/gmerp_comments/enstrom.pdf). The March 23, 2006 CARB meeting PPT presentation “Stronger Relationship Between Particulate Matter (PM) and Premature Death” gave extensive details on Jerrett 2005 and cited several other positive national studies, including Krewski 2000, Pope 2002, and Laden 2006 (<http://www.arb.ca.gov/research/health/healthup/march06.pdf>). However, it made no mention of Enstrom 2005, which was published one month after Jerrett 2005 and one month before a major Harvard Six Cities Study analysis (Laden 2006) appeared online. On August 21, 2006 CARB scientists Richard Bode, Linda Smith, and Hien T. Tran conducted a “Public Workshop on Updating the Methodology for Estimating Premature Death Associated with PM_{2.5} Exposures” and gave a PPT presentation (<http://www.arb.ca.gov/research/health/pm-mort/ws-slides.pdf>). The PPT presentation for this Workshop specifically shows Jerrett 2005 and Laden 2006, but not Enstrom 2005, as “New studies emerged since 2002.” These PPT presentations show a pattern of omission of null findings like Enstrom 2005.

Additional misrepresentation of PM_{2.5} mortality risk in California was contained in the Draft and Final versions of the 2008 CARB Staff Report by Hien T. Tran “Methodology for Estimating Premature Deaths Associated with Long-term Exposure to Fine Airborne Particulate Matter in California.” The October 24, 2008 Final Report states that PM_{2.5} contributes to 18,000 annual premature deaths in California, with 3,500 of these deaths due to diesel PM. These estimates of premature deaths provided the primary public health justification for new on-road diesel vehicle regulations approved and implemented by CARB. However, the premature death claims in this report are now entirely contradicted by the null findings presented in Table 1. My December 10, 2008 CARB comments exposed major flaws in this report (http://www.arb.ca.gov/lists/truckbus08/897-carb_enstrom_comments_on_statewide_truck_regulations_121008.pdf). The CARB misrepresentations of PM_{2.5} mortality risk in California continue up to the present, as explained in my talks and submissions cited above.

2.3 Failure to Properly Review Particulate Matter Health Impacts by AQMD

As an essential part of its currently ongoing preparation of the 2012 AQMP, the AQMD is required to address the health effects of air pollution in the SCAB. Indeed, California Health and Safety Code (CHSC) Section 40471 (b) specifically states “On or before December 31, 2001, and every three years thereafter, as part of the preparation of the air quality management plan revisions, the south coast district board, in conjunction with a public health organization or agency, shall prepare a report on the health impacts of

particulate matter air pollution in the South Coast Air Basin. The south coast district board shall submit its report to the advisory council appointed pursuant to Section 40428 for review and comment. The advisory council shall undertake peer review concerning the report prior to its finalization and public release. The south coast district board shall hold public hearings concerning the report and the peer review, and shall append to the report any additional material or information that results from the peer review and public hearings.” (<http://www.leginfo.ca.gov/cgi-bin/displaycode?section=hsc&group=40001-41000&file=40460-40471>).

However, based on available information, AQMD has never prepared a “report on the health impacts of particulate matter air pollution in the South Coast Air Basin” at the end of 2001, 2004, 2007, or 2010. The only “health impacts” reports are Appendix I “Health Effects” of the 2003 AQMP, 2007 AQMP, and Draft 2012 AQMP. However these reports do not specifically address PM health impacts in the SCAB. Indeed, the 2003 AQMP Appendix I states “The purpose of this appendix is to provide an overview of air pollution health effects, rather than to provide estimates of health risk from current ambient levels of pollutants in specific areas of the SCAB.” (http://www.aqmd.gov/aqmp/docs/2003AQMP_AppI.pdf).

Failure to comply with CHSC Section 40471 (b) is a serious matter because the local health effects of PM provide the primary public health justification for the entire AQMP. As shown in Tables 1 and 2, there is now overwhelming epidemiologic evidence that there is NO relationship in California between PM and total mortality (also known as "premature deaths"). However, the 2003 AQMP Appendix I (https://aqmd.gov/aqmp/docs/2003AQMP_AppI.pdf, page I-14), 2007 AQMP Appendix I (https://aqmd.gov/aqmp/07aqmp/aqmp/Appendix_I.pdf, page I-14), 2012 Draft AQMP Appendix I (<http://www.aqmd.gov/aqmp/2012aqmp/draft/Appendices/AppxI.pdf>, page I-18), and 2012 Revised Draft AQMP Appendix I (<http://www.aqmd.gov/aqmp/2012aqmp/RevisedDraft/AppI.pdf>, page I-19) all make incorrect statements regarding the evidence in California and the SCAB.

All four Health Effects appendices have been authored by AQMD Health Effects Officer Jean Ospital (http://www.aqmd.gov/bios/ms_ospital_jean.html). These documents come to exactly the same conclusion regarding PM mortality risk: “Despite data gaps, the extensive body of epidemiological studies has both qualitative and quantitative consistency suggestive of causality. A considerable body of evidence from these studies suggests that ambient particulate matter, alone or in combination with other coexisting pollutants, is associated with significant increases in mortality and morbidity in a community. In summary, the scientific literature indicates that an increased risk of mortality and morbidity is associated with particulate matter at ambient levels. The evidence for particulate matter effects is mostly derived from population studies with supportive evidence from clinical and animal studies.”

The null PM_{2.5} - mortality relationship in California has been known since 2000, but the specific null evidence is only partially presented in the Draft 2012 AQMP and was entirely omitted from the earlier AQMPs. For instance, each AQMP Appendix I cites Krewski 2000. However, only the nationwide PM_{2.5} mortality risk results in this report are cited, not the California-specific results in Figure 21. The 2007 AQMP Appendix review cites Jerrett 2005, Laden 2006, and the Pope 2006 review, which contains two references to Enstrom 2005, but Enstrom 2005 itself is not mentioned. Enstrom 2005 is mentioned briefly in the Draft 2012 Appendix I, but not assigned any major significance.

The overwhelmingly null evidence in Figures 1 and 2 is not fully or properly described in either the Draft or Revised Draft 2012 Appendix I. I pointed out major deficiencies in my April 21, 2011 CARB comments (http://www.arb.ca.gov/lists/sip2011/3-carb_enstrom_comments_on_sip_for_pm2.5_042711.pdf). Since August 2008 I have also had repeated direction communications with Ospital, including an April 4, 2012 email message requesting that null evidence be included in the 2012 AQMP Appendix I (<http://www.scientificintegrityinstitute.org/Ospital040412.pdf>).

The health impacts of PM in the SCAB are still not addressed in the September 7, 2012 Revised 2012 Draft AQMP Appendix I (<http://www.aqmd.gov/aqmp/2012aqmp/RevisedDraft/AppI.pdf>). Furthermore, this version makes an incorrect assessment of the California-specific evidence by uncritically relying on the June 2012 US EPA Regulatory Impact Analysis (RIA) (US EPA 2012). The RIA looked at California-specific studies regarding PM_{2.5} and mortality published in the scientific literature. Appendix I states “The EPA analysis concluded ‘most of the cohort studies conducted in California report central effect estimates similar to the (nation-wide) all-cause mortality risk estimate we applied from Krewski et al. (2009) and Laden et al. (2006) albeit with wider confidence intervals. A couple cohort studies conducted in California indicate higher risks than the risk estimates we applied.’ Thus in EPAs judgment the California related studies provided estimates of mortality consistent with or higher than those from the national studies.”

However, there are clear errors in virtually every California-specific RR in EPA RIA Table 5.B-10. The McDonnell 2000 ratio, RR (males) = 1.09 (0.98–1.24), should be RR (both sexes) ~ 1.00 (0.95–1.05), based on inclusion of an approximated RR for females. The partially adjusted Jerrett 2005 ratio, RR = 1.15 (1.03–1.29), should be the fully adjusted value, RR = 1.11 (0.99–1.25). The Enstrom 2005 ratio for 1973-1982, RR = 1.04 (1.01–1.07), should be the ratio for the entire follow-up period (1973-2002), RR = 1.01 (0.99–1.03). The Krewski 2009 ratio, RR = 1.42 (1.26–1.27), is obviously invalid and should be replaced by the Krewski 2010 ratio, RR = 0.968 (0.916–1.022), which is the ratio for all California subjects in Krewski 2009. The implausibly high Ostro 2010 ratio, RR = 1.84 (1.66–2.05), is invalid and has been replaced by the new Ostro 2011 ratio, RR = 1.06 (0.96–1.16). The corrected ratios are all consistent with RR = 1.00 and DO NOT support the EPA RIA claim that California-specific results are consistent with national results. Ospital uncritically accepted the EPA RIA and did not mention a single one of the EPA errors cited above.

The July 11, 2012 AQMP Advisory Council meeting did not result in proper peer review of Draft 2012 Appendix I. The three Advisory Council members with the most expertise on PM mortality studies and PM health effects epidemiology are John R. Froines, Ph.D., Samuel Soret, Ph.D., and Rob S. McConnell, M.D. They have not done peer review of Appendix I regarding “the health impacts of particulate matter air pollution in the South Coast Air Basin,” as specified in CHSC Section 40471 (b). Also, there is evidence that they are not objective peer reviewers regarding PM health effects.

UCLA Professor John R. Froines has engaged in inappropriate activism regarding PM science based on the information contained in the following documents:

- 1) June 30, 2009 letter and attachments from Norman R. Brown to UCLA officials (http://www.calcontrk.org/CARBdocs/Delta_UCLA_Letter_063009.pdf),
- 2) February 20, 2011 Bakersfield Californian column by Lois Henry

(<http://www.bakersfieldcalifornian.com/columnists/lois-henry/x1902890284/Politics-air-rules-make-for-a-smelly-situation>), and 3) April 15, 2012 Bakersfield Californian column by Lois Henry (<http://www.bakersfieldcalifornian.com/health/x1322083219/The-ex-radical-who-heads-air-boards-key-panel>).

Loma Linda University (LLU) Professor Samuel Soret has not responded to my August 23, 2012 and September 14, 2012 email messages regarding his peer review of the AQMP Appendix I (<http://www.scientificintegrityinstitute.org/Soret091412.pdf>). His July 11, 2012 email message to AQMD did not mention the highly relevant December 2010 paper that he co-authored and apparently submitted to *Epidemiology* "The Mortality & Long-Term Exposure to AP in Elderly CA Adventists" (Chen 2010). Also, he has not properly described the overwhelmingly null relationship between PM and total mortality in the 35-year LLU Adventist Health Study of Air Pollution (AHSMOG) project (<http://www.llu.edu/public-health/health/ahsmog.page>).

USC Professor Rob S. McConnell has not responded to my August 25, 2012 and September 17, 2012 email messages regarding his incomplete July 9, 2012 peer review of AQMP Appendix I, which did not discuss PM in the SCAB (<http://www.scientificintegrityinstitute.org/McConnell091712.pdf>).

I submitted comments to AQMD regarding AQMP Appendix I on August 30, 2012 (<http://www.scientificintegrityinstitute.org/AQMP083012.pdf>) and on September 20, 2012 (<http://www.scientificintegrityinstitute.org/AQMP092012.pdf>).

These comments emphasize the need for AQMD to comply with all provisions of CHSC Section 40471 (b) before finalizing the 2012 AQMP. It is particularly important that the AQMD Governing Board conduct a hearing on the health impacts of PM in the SCAB. This hearing will allow scientists with diverse views to directly present evidence to the Board Members. This hearing could have a profound impact on the emission control measures that are approved in the 2012 AQMP.

Conclusions

There is now overwhelming epidemiologic evidence that PM (PM_{2.5} and PM₁₀) is not killing Californians. This evidence must be fully examined and recognized by EPA, CARB, and AQMD before there are any further regulations to reduce PM levels in California, particularly in the SCAB. In addition, there needs to be a full reassessment of the current PM regulations to be sure that they are based on the actual health effects evidence in California. AQMD should not be required to comply with NAAQS that are not appropriate for California or the SCAB. Instead, AQMD should request a waiver from compliance with the NAAQS using the special waiver status granted to California in Section 209 of the Clean Air Act (<http://www.epa.gov/otaq/cafr.htm>). Finally, PM health effects and regulations must be put into perspective with other factors that influence health in California. Keep in mind the findings in Figure 2, which show that, based on the 2009 age-adjusted total death rate by state, California had the third lowest rate. Furthermore, the SCAB had a total death rate that was lower than the rate for every state except Hawaii (<http://www.scientificintegrityinstitute.org/NCHSRR070811.pdf>).

Table 1. Epidemiologic Cohort Studies of PM_{2.5} and Total Mortality in California
<http://www.scientificintegrityinstitute.org/Enstrom081512.pdf>

Relative risk of death from all causes (RR and 95% CI) associated with increase of 10 µg/m³ in PM_{2.5}

Krewski 2000 & 2010	CA CPS II Cohort (N=40,408 [18,000 M + 22,408 F]; 4 MSAs; 1979-1983 PM _{2.5} ; 44 covariates)	RR = 0.872 (0.805-0.944)	1982-1989
McDonnell 2000	CA AHSMOG Cohort (N~3,800 [1,347 M + 2,422 F]; SC&SD&SF AB; M RR=1.09(0.98-1.21) & F RR~0.98(0.92-1.03))	RR ~ 1.00 (0.95 – 1.05)	1977-1992
Jerrett 2005	CPS II Cohort in Los Angeles Basin (N=22,905; 267 zip code areas; 1999-2000 PM _{2.5} ; 44 cov + max confounders)	RR = 1.11 (0.99 - 1.25)	1982-2000
Enstrom 2005	CA CPS I Cohort (N=35,783 [15,573 M + 20,210 F]; 11 counties; 1979-1983 PM _{2.5} ; 25 county internal comparison)	RR = 1.039 (1.010-1.069) RR = 0.997 (0.978-1.016)	1973-1982 1983-2002
Enstrom 2006	CA CPS I Cohort (N=35,783 [15,573 M + 20,210 F]; 11 counties; 1979-1983 & 1999-2001 PM _{2.5})	RR = 1.061 (1.017-1.106) RR = 0.995 (0.968-1.024)	1973-1982 1983-2002
Zeger 2008	MCAPS Cohort “West” (3.1 M [1.5 M M + 1.6 M F]; Medicare enrollees in CA+OR+WA (CA=73%); 2000-2005 PM _{2.5})	RR = 0.989 (0.970-1.008)	2000-2005
Jerrett 2010	CA CPS II Cohort (N=77,767 [34,367 M + 43,400 F]; 54 counties; 2000 PM _{2.5} ; KRG ZIP; 20 ind cov+7 eco var; Slide 12)	RR ~ 0.994 (0.965-1.025)	1982-2000
Krewski 2010	CA CPS II Cohort (N=40,408; 4 MSAs; 1979-1983 PM _{2.5} ; 44 cov) (N=50,930; 7 MSAs; 1999-2000 PM _{2.5} ; 44 cov)	RR = 0.960 (0.920-1.002) RR = 0.968 (0.916-1.022)	1982-2000 1982-2000
Jerrett 2011	CA CPS II Cohort (N=73,609 [32,509 M + 41,100 F]; 54 counties; 2000 PM _{2.5} ; KRG ZIP Model; 20 ind cov+7 eco var; Table 28)	RR = 0.994 (0.965-1.024)	1982-2000
Jerrett 2011	CA CPS II Cohort (N=73,609 [32,509 M + 41,100 F]; 54 counties; 2000 PM _{2.5} ; Nine Model Ave; 20 ic+7 ev; Fig 22 & Tab 27-32)	RR = 1.002 (0.992-1.012)	1982-2000
Lipsett 2011	CA Teachers Cohort (N=73,489 [73,489 F]; 2000-2005 PM _{2.5})	RR = 1.01 (0.95 – 1.09)	2000-2005
Ostro 2011	CA Teachers Cohort (N=43,220 [43,220 F]; 2002-2007 PM _{2.5}) replaced Ostro 2010	RR = 1.06 (0.96 – 1.16) Incorrect 2010 Result: RR = 1.84 (1.66 – 2.05)	2002-2007 2002-2007

Table 2. Epidemiologic Cohort Studies of PM10 and Total Mortality in California

Relative risk of death from all causes (RR and 95% CI) associated with increase of 10 $\mu\text{g}/\text{m}^3$ in PM10

Abbey 1999	CA AHSMOG Cohort	M	RR = 1.04	(0.99 – 1.10)	1977-1992
(N=6,338 [2,278 M + 4,060 F]; PM10		F	RR = 0.98	(0.93 – 1.02)	1977-1992
monitors in SC & SD & SF Air Basins)		BS	RR = 1.00	(0.97 – 1.04)	1977-1992
[N=610M+965F, all natural causes ICD9=001-799]					
McDonnell 2000	CA AHSMOG Cohort	M	RR = 1.05	(0.98 – 1.12)	1977-1992
(N~3,800 [1,347 M + 2,422 F]; PM10		F	RR ~ 0.98	(0.92 – 1.03)	1977-1992
monitors in SC & SD & SF Air Basins)		BS	RR ~ 1.01	(0.96 – 1.05)	1977-1992
[all natural causes ICD9=001-799]					
Chen 2010	CA AHSMOG Cohort		RR = 1.01	(0.98 – 1.04)	1977-2006
(N=4,830 [1,750 M + 3,080 F]; PM10					
monitors in SC & SD & SF Air Basins)					
[all natural causes ICD9=001-799]					
Jerrett 2011	CA CPS II Cohort		RR = 1.001	(0.987-1.017)	1982-2000
(N=76,135 [33,625 M + 42,510 F]; 54 counties;					
1988-2002 PM10; KRG Zip Model; 20 ind cov+7 eco var; Table 37)					
Lipsett 2011	CA Teachers Cohort		RR = 1.01	(0.95 – 1.09)	2000-2005
(N=73,489 [73,489 F]; 2000-2005 PM10)					

Figure 1. Figures 21 and 5 from HEI Reanalysis Report (Krewski 2000)

Figure 21 Spatial Overlay of PM2.5 Level and Mortality Risk by City (page 197)
Fine Particles and Mortality Risk

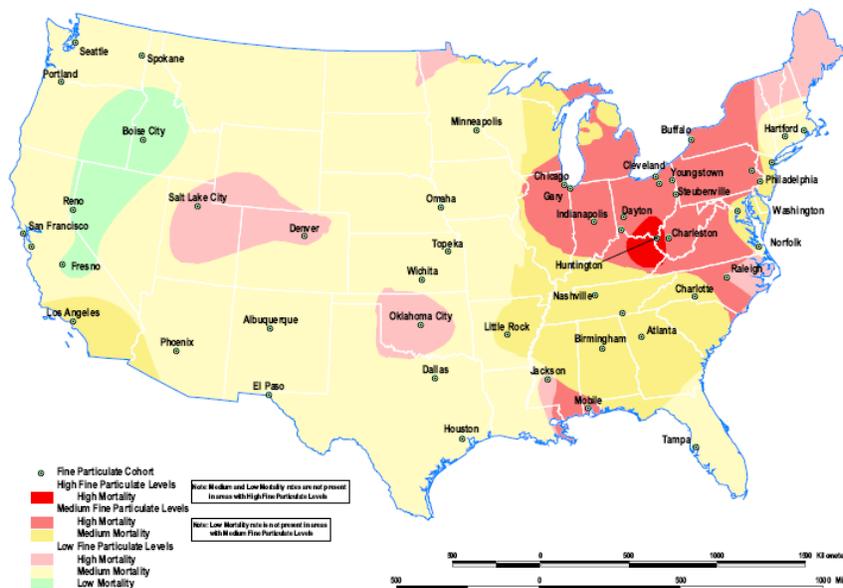


Figure 21. Spatial overlay of fine particle levels and relative risk of mortality. Interval classifications for fine particles ($\mu\text{g}/\text{m}^3$): low 8.99–17.03; medium 17.03–25.07; high 25.07–33.07. Interval classifications for relative risks of mortality: low 0.502–0.711; medium 0.711–0.919; high 0.919–1.128.

107

Figure 5 (Upper Right) Relative Risk for PM2.5 and Total Mortality by City (page 161)
All Cause Mortality (Excluding Boise City, Idaho)

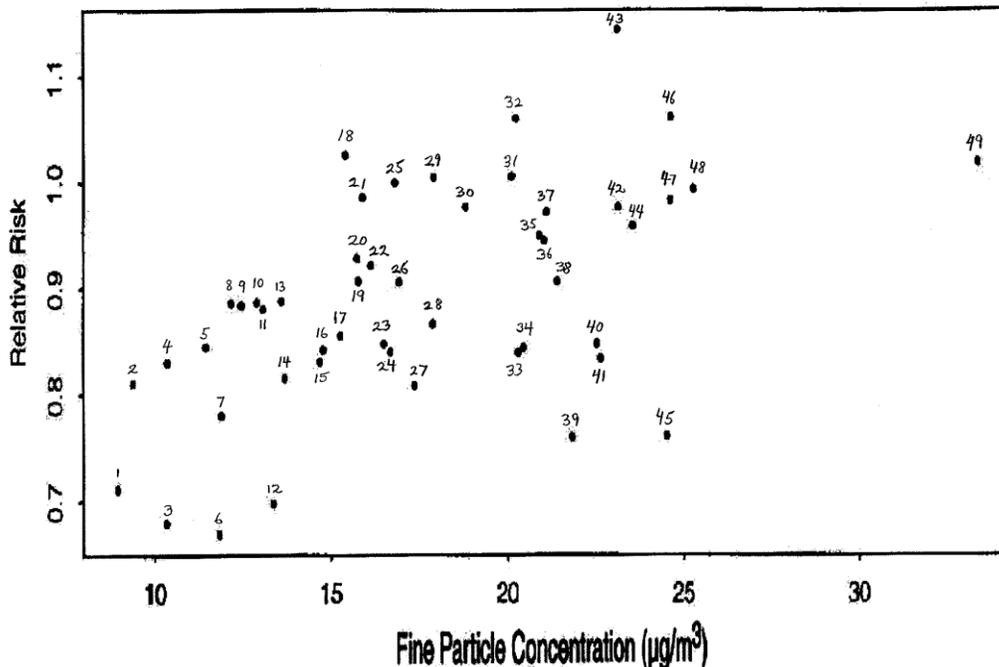
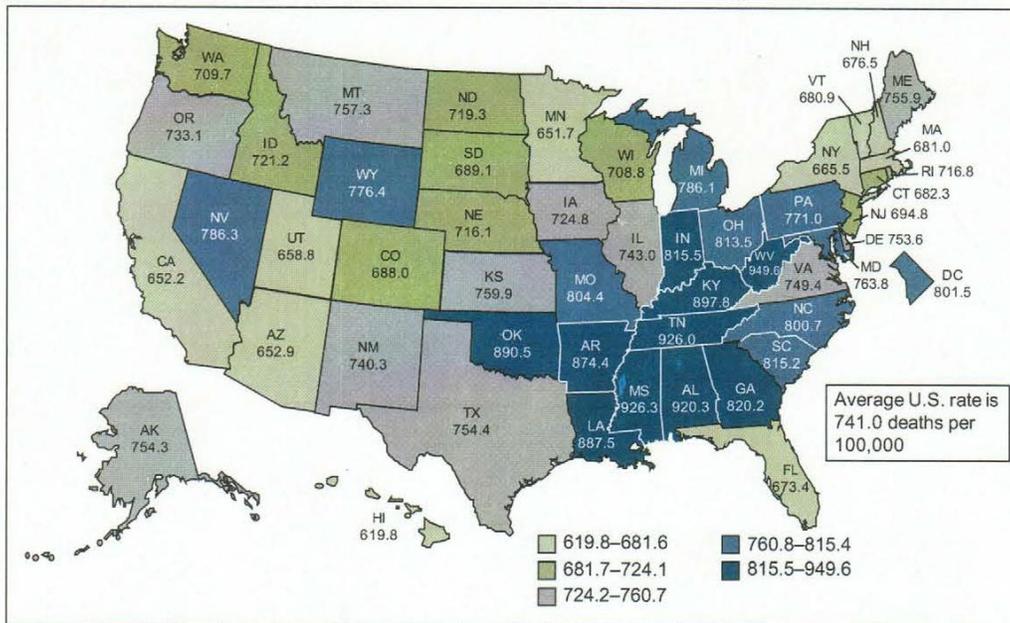


Figure 2. 2009 Age-Adjusted Total Death Rates by State for the United States
 NCHS Data Brief Number 64, July 2011 “Death in the United States, 2009”
<http://www.cdc.gov/nchs/data/databriefs/db64.pdf>
<http://www.scientificintegrityinstitute.org/NCHSDR070811.pdf>

Figure 4. Age-adjusted death rates, by state and the District of Columbia: United States, preliminary 2009



SOURCE: CDC/NCHS, National Vital Statistics System, Mortality.

Ratio of 2009 Age-Adjusted Total Death Rates (Deaths/100,000)

California / U.S.	$652.2 / 741.1 = 0.880 = 88.0\%$
‘South Coast Air Basin’ (4 Counties) / U.S.	$650.8 / 741.1 = 0.878 = 87.8\%$
Los Angeles County / U.S.	$637.3 / 741.1 = 0.860 = 86.0\%$
Orange County / U.S.	$570.9 / 741.1 = 0.770 = 77.0\%$

References

Abbey DE, Nishino N, McDonnell WF, Burchette RJ, Knutsen SF, Lawrence Beeson W, Yang JX (1999). Long-term inhalable particles and other air pollutants related to mortality in nonsmokers. *Am J Respir Crit Care Med* 1999;159:373-382
<http://ajrccm.atsjournals.org/content/159/2/373.full.pdf>

Chen LH, Knutsen SF, Beeson WL, Soret S, Shavlik D, Ghamsary M (2010). “The Mortality and Long-Term Exposure to Ambient Air Pollution in Nonsmoking Adults” (<http://www.scientificintegrityinstitute.org/Chen4C2010.pdf>), Chapter 4 of December 2010 Loma Linda University Dr.P.H. Dissertation "Coronary Heart Disease Mortality and Long-term Exposure to Ambient Particulate Air Pollutants in Elderly Nonsmoking California Residents" by Lie H. Chen (<http://www.scientificintegrityinstitute.org/Chen2010.pdf>)

Enstrom JE (2005). Fine particulate air pollution and total mortality among elderly Californians, 1973-2002. *Inhal Toxicol* 2005;17:803-816 (http://www.arb.ca.gov/planning/gmerp/dec1plan/gmerp_comments/enstrom.pdf) and (<http://www.scientificintegrityinstitute.org/IT121505.pdf>)

Enstrom JE (2006). Response to “A Critique of 'Fine Particulate Air Pollution and Total Mortality Among Elderly Californians, 1973-2002’ by Bert Brunekreef, PhD, and Gerard Hoek, PhD’, *Inhal Toxicol* 2006;18:509-514 (<http://www.scientificintegrityinstitute.org/IT060106.pdf>) and (<http://www.scientificintegrityinstitute.org/ITBH060106.pdf>)

Jerrett M, Burnett RT, Ma R, Pope CA III, Krewski D, Newbold KB, Thurston G, Shi Y, Finkelstein N, Calle EE, Thun MJ (2005). Spatial Analysis of Air Pollution and Mortality in Los Angeles. *Epidemiology* 2005;16:727–736 (<http://www.scientificintegrityinstitute.org/Jerrett110105.pdf>)

Jerrett M (2010). February 26, 2010 CARB Symposium Presentation by Principal Investigator, Michael Jerrett, UC Berkeley/CARB Proposal No. 2624-254 "Spatiotemporal Analysis of Air Pollution and Mortality in California Based on the American Cancer Society Cohort” (<http://www.scientificintegrityinstitute.org/CARBJerrett022610.pdf>)

Jerrett M (2011). October 28, 2011 Revised Final Report for Contract No. 06-332 to CARB Research Screening Committee, Principal Investigator Michael Jerrett, “Spatiotemporal Analysis of Air Pollution and Mortality in California Based on the American Cancer Society Cohort” Co-Investigators: Burnett RT, Pope CA III, Krewski D, Thurston G, Christakos G, Hughes E, Ross Z, Shi Y, Thun M (<http://www.arb.ca.gov/research/rsc/10-28-11/item1dfr06-332.pdf>) and (<http://www.scientificintegrityinstitute.org/Jerrett012510.pdf>) and (<http://www.scientificintegrityinstitute.org/JerrettCriticism102811.pdf>)

Krewski D (2000). "Reanalysis of the Harvard Six Cities Study and the American Cancer Society Study of Particulate Air Pollution and Mortality: HEI Special Report. July 2000" (<http://pubs.healtheffects.org/view.php?id=6>). Figure 5 and Figure 21 of Part II: Sensitivity Analyses (<http://pubs.healtheffects.org/getfile.php?u=275>) and (<http://www.scientificintegrityinstitute.org/HEIFigure5093010.pdf>)

Krewski D (2009). Krewski D, Jerrett M, Burnett RT, Ma R, Hughes E, Shi Y, Turner MC, Pope CA III, Thurston G, Calle EE, Thun MJ. [Extended Analysis of the American Cancer Society Study of Particulate Air Pollution and Mortality](#). HEI Research Report 140. May 2009 (<http://pubs.healtheffects.org/view.php?id=315>)

Krewski D (2010). August 31, 2010 letter from Krewski to Health Effects Institute and CARB with California-specific PM2.5 mortality results from Table 33 in Krewski 2009 (http://www.arb.ca.gov/research/health/pm-mort/HEI_Correspondence.pdf)

Laden F, Schwartz J, Speizer FE, Dockery DW (2006). Reduction in fine particulate air pollution and mortality: Extended follow-up of the Harvard Six Cities Study. *Am J Respir Crit Care Med* 2006;173:667-672. Epub 2006 Jan 19 (<http://ajrccm.atsjournals.org/content/173/6/667.full.pdf>)

Lipsett MJ, Ostro BD, Reynolds P, Goldberg D, Hertz A, Jerrett M, Smith DF, Garcia C, Chang ET, Bernstein L (2011). Long-term Exposure to Air Pollution and Cardiorespiratory Disease in the California Teachers Study Cohort. *Am J Respir Crit Care Med* 2011;184:828–835 (<http://ajrccm.atsjournals.org/content/184/7/828.full.pdf>)

McDonnell WF; Nishino-Ishikawa N; Petersen FF; Chen LH; Abbey DE (2000). Relationships of mortality with the fine and coarse fractions of long-term ambient PM10 concentrations in nonsmokers. *J Expo Anal Environ Epidemiol* 2000;10:427-436 (<http://www.scientificintegrityinstitute.org/JEAAE090100.pdf>)

Moolgavkar SH (2006). Fine Particles and Mortality. *Inhal Toxicol* 2006;18:93-94 (<http://www.scientificintegrityinstitute.org/IT010106.pdf>)

Ostro B, Lipsett M, Reynolds P, Goldberg D, Hertz A, Garcia C, Henderson KD, Bernstein L (2010). Long-Term Exposure to Constituents of Fine Particulate Air Pollution and Mortality: Results from the California Teachers Study. *Environ Health Perspect* 2010;118:363-369 with June 2011 Erratum (<http://ehp03.niehs.nih.gov/article/info:doi/10.1289/ehp.0901181>)

Pope CA III, Burnett RT, Thun MJ, Calle EE, Krewski D, Ito K, Thurston GD (2002). Lung cancer, cardiopulmonary mortality, and long-term exposure to fine particulate air pollution. *JAMA* 2002;287:1132-1141 (<http://jama.jamanetwork.com/data/Journals/JAMA/4822/JOC11435.pdf>)

Pope CA III, Dockery DW. (2006). Health Effects of Fine Particulate Air Pollution: Lines that Connect. *JAWMA*, Critical Review. 56(6):709-742 (<http://www.scientificintegrityinstitute.org/PopeDockery2006.pdf>) and (<http://www.scientificintegrityinstitute.org/PopePPT2006.pdf>)

U.S. EPA (2012). Regulatory Impact Analysis related to the Proposed Revisions to the National Ambient Air Quality Standards for Particulate Matter EPA-452/R-12-003 (http://www.epa.gov/ttn/ecas/regdata/RIAs/PMRIACombinedFile_Bookmarked.pdf)

Zeger SL, Dominici F, McDermott A, Samet JM (2008). Mortality in the Medicare Population and Chronic Exposure to Fine Particulate Air Pollution in Urban Centers (2000-2005). *Environ Health Perspect* 2008;116:1614-1619 (<http://ehp03.niehs.nih.gov/article/info:doi/10.1289/ehp.11449>)

John Dale Dunn MD JD
Diplomate ABEM, ABLM
Admitted but inactive, Texas and Louisiana Bars
Consultant Emergency Services, Peer Review
Mediator

401 Rocky Hill Road

Lake Brownwood, Texas 76801

Phone 325 784-6697
Fax 325 784-7567, call first
E-mail jddmdjd@web-access.net

June 8, 2011

Research Screening Committee Members

California Air Resources Board
1001 I Street
P.O. Box 2815
Sacramento, CA 95812

RE: Draft report for the contract No. 06-332 "Spatiotemporal Analysis of Air Pollution and Mortality in California Based on the American Cancer Society Cohort"

Ladies and Gentlemen of the Screening Committee,

Your have a choice in your consideration of this study by Dr. Michael Jerrett and many Co-Authors on whether you will properly execute your duties to assure good science informs good policy making, or you can be complicit in a scientific fraud of great magnitude. This study and report, particularly its conclusions, are a scientific fraud that not only ignores the rules of epidemiology and good human health effects science, but are complicit in fraudulent activity that uses public moneys, by faculty members of the University of California and others who put their names to the study.

I have reviewed the "Jarrett" study, paid for by 750,000 taxpayer dollars, which is an important consideration expanded on herein below. The Jarrett 3 year effort is based on assumptions that are derivative of previous studies, but in the main it is a modeling exercise intended to dredge for proof that there are small particle air pollution deaths that justify a California Air Resources Board small particle regulatory regime. Nothing in this expensive desk top computer modeling study is adequate to the task. After all is said and done, now looking at the Jarrett study, it shows no evidence that current ambient small particles in the air of California air are killing anyone.

Here is where the fraud begins, members of the Screening Committee.

The models failed to provide the proof that Dr. Enstrom was wrong in 2005 when he said there is no small particle death effect in California. The elaborate Jarrett study confirms what Jarrett admitted in February of 2010, that he could find no human health effect from California small particle air pollution. The study presented to the committee fails to disprove or contradict the assertion of Dr. Enstrom in 2005 or the admission of Dr. Jarrett in 2010 that CARB claims of deaths from small particles were not evident in his research. Dr. Jarrett in 2010 was admitting that, even as the chosen researcher for CARB, he could not find evidence to show

death effects from small particles in the air.

The only model in the elaborate and thick Jarrett study before you that provides even a glimmer, A GLIMMER, for the CARB agenda of small particle regulations failed when the minor relative risk of 1.08 was combined with a confidence interval that included 1.0. ATTENTION, LADIES AND GENTLEMEN OF THE SCREENING COMMITTEE—THAT MEANS THAT THE JARRETT STUDY SHOWS NO SMALL PARTICLE EFFECTS. PERIOD. NONE, IN ANY OF THE MODELS OR ALTERNATIVE SCENARIOS.

However, because this is such a scandal, and because criticizing Dr. Jarrett's study is so easy, I would like to list a few points for your consideration: ,

1. The Jarrett study, if intended to show small particles kill, came a cropper (that means it failed, folks), since it fails in every effort to find significant evidence that small particles kill Californians. In fact it shows what we all knew, that Californians are not dying from small particles. All of the studies showed effects with a confidence interval that crossed or included 1.0. As Bugs Bunny would say—that's all folks! You have nothing to hang your hat on and approval of this study will show your lack of good faith.
2. All 9 modeling exercises, intended to dredge for proof to support CARB had no effects that escaped the confidence interval that made them mean nothing—NOTHING. The studies showed the confidence intervals meeting or crossing 1.0, confirming that there is of NO EFFECT of small particles on premature death in California from small particles of 2.5 microns or less.
3. When the 9 studies offered by the Jarrett study show no effect, any CARB decision to pursue the Small Particle regulations would not only violate a committee public duty to pursue policies that are based on sound science, I WOULD ARGUE THAT SUCH A DECISION BY CARB WOULD INDICATE COMPLICITY BY THE COMMITTEE AND BY CARB LEADERSHIP IN A FRAUD, A FRAUDULENT STUDY PAID FOR BY THE BELEAGUED TAXPAYERS OF CALIFORNIA WHO COULD HAVE BEEN SPARED THE THREE QUARTERS OF A MILLION DOLLARS WASTED ON THE STUDY.
4. I would remind the review committee that complicity in a fraud exposes individuals, either in their official or their individual capacities as parties to misuse of taxpayer funds.

I will not belabor the members of the committee with the epidemiological rules and the toxicology rules that are applicable to studies such as the Jarrett study. Suffice it to say that Federal Judicial Rules of Evidence specify that scientific evidence such as that contained in the Jarrett study should be reliable and relevant for the case in hand—the question of whether CARB has the science to justify its policy decisions.

The misrepresentation and fraud of the Jarrett group and the Jarrett study is most evident in the conclusions. The authors state “We conclude that combustion-source air pollution, especially from traffic, is significantly associated with premature death in this large cohort of Californians.” A reasonable citizen reviewer of the study, knowledgeable in the science of epidemiology would ask--how could the authors use words like “conclude” or “significantly associated” when they have nothing in the study to support an assertion?

Have the authors sold their scientific integrity for \$750,000? Are they implicated in a fraud on the citizens of California, claiming their “show nothing” study is adequate to support a new ambitious and onerous CARB regulatory regime focused on small particles?

There is retribution in the law for fraud on the taxpayers. Laws were enacted to prevent dishonest and

fraudulent use of public moneys. Committees that fail to recognize their responsibility as fiduciaries for the taxpayers could also be considered complicit in the fraud if they have been properly warned.

This letter is proper warning to the members of the review committee.

Consider your options when I am telling you, as an experienced and knowledgeable man of science and the law. You and the CARB and the scientists involved in this disgraceful study may have to answer questions on whether the study was properly conducted, but more importantly, were the conclusions proper, given the evidence or, were those conclusions bought and paid for?

Respectfully,

John Dale Dunn MD JD

Particulate Matter Does Not Cause Premature Deaths

James E. Enstrom, Ph.D., M.P.H., FACE, AAAS 40-year Member
UCLA and Scientific Integrity Institute

S. Stanley Young, Ph.D., FASA, FAAAS
Raleigh, NC

John Dale Dunn, M.D., J.D.
JunkScience.com

Charles G. Battig, M.D., M.S.
VA Scientists & Engineers for Energy & Environment

William M. Briggs, Ph.D.
wmbriggs.com

Edward J. Calabrese, Ph.D.
University of Massachusetts, Amherst, MA

Alan Carlin, Ph.D.
CarlinEconomics.com

Laurence I. Gould, Ph.D.
University of Hartford, CT

Peter W. Wood, Ph.D., AAAS Member
National Association of Scholars

August 17, 2015

An extensive 2011 U.S. Environmental Protection Agency (EPA) cost-benefit report estimates the annual costs required to meet 1990 Clean Air Act (CAA) Amendment regulations to be about \$65 billion in 2020. The annual economic benefits of these regulations are estimated to be about \$2 trillion in 2020, based primarily on EPA-projected reductions in air pollution-related premature deaths and illness (1). This report has been challenged because the benefits are unproven and depend upon several questionable and unverified assumptions. Among these are assumptions that a linear, no-threshold, *causal* relation exists between fine particulate air pollution (PM_{2.5}) and total mortality and that additional life expectancy gained at a median age of about 80 years should be valued at about \$80,000 per month. These assumptions are essential because \$1.7 trillion (85%) of the \$2.0 trillion total benefit estimate is attributable to reductions in premature deaths due to reductions in PM_{2.5}. Using discrete uncertainty analysis with plausible alternative assumptions, Cox found that the costs of CAA amendments actually exceed their benefits (2).

Dominici et al. have stated: “With the estimated benefits of PM reductions playing such a central role in regulatory policy, it is critical to ensure that the estimated health benefits are based on the best available evidence. If the estimates are biased upward (downward), then the regulations may be too stringent (lenient).” (3). Because of the urgent need to verify the health benefits of EPA regulations, Congress is enacting the Secret Science Reform Act (SSRA) (4). The SSRA would “prohibit the Environmental Protection Agency from proposing, finalizing, or disseminating regulations or assessments based upon science that is not transparent or reproducible.”

Based on the data and research findings that are currently available without the SSRA, we challenge the validity of the annual \$1.7 trillion health benefit attributed to reductions in PM_{2.5}. Specifically, we present four types of evidence that PM_{2.5} does not *cause* premature deaths.

1) The major increase in U.S. life expectancy since 1970 is not due to reduction in PM_{2.5}. In 2009 Pope claimed that from 1980 to 2000 a decrease of 10 µg/m³ of PM_{2.5} was associated nationally with a 0.61 year increase in life expectancy based on a correlation involving 51 U.S. metropolitan areas (USMAs) (5). This association was vigorously contested by four independent analyses because the underlying data was available, as would be required by the SSRA. Enstrom found no association whatsoever in 11 California counties (5). Krstic found that the national association claimed by Pope lost statistical significance with the removal of one USMA (Topeka, KS) and that the correlation between changes in PM_{2.5} and life expectancy had so much scatter that it explained almost none of the association (6). Young showed that there was no association in the Western U.S., thereby supporting Enstrom, and showed that the national association was much stronger with income than with PM_{2.5} (7). Cox found no significant association between reductions in PM_{2.5} and total mortality rate between 2000 and 2010 in 483 counties in the 15 most populated states, including California (8). The inconsistencies and weaknesses found in the association means that Pope did not prove the hypothesis that a reduction in PM_{2.5} *causes* an increase in life expectancy. However, since 1970, the year that EPA was established, health-related factors other than air pollution have had a major impact on increasing the longevity of Americans. The total annual age-adjusted death rate in the U.S. has declined by 40% from 12.226 deaths/1000 in 1970 to 7.319 deaths/1000 in 2013. The death rate in California has declined by 45% from 11.370 deaths/1000 in 1970 to 6.301 deaths/1000 in 2013. Life expectancy from birth has increased from 70.8 years in 1970 to 78.8 years in 2013 in the U.S. and from 71.7 years in 1970 to 80.8 years in 2013 in California (9).

2) No plausible etiologic mechanism by which PM_{2.5} *causes* premature death is established. It is implausible that a never-smoker’s death could be *caused* by inhalation over an 80 year lifespan of about one teaspoon (~5 grams) of invisible fine particles as a result of daily exposure to 15 µg/m³. This level of exposure is equivalent to smoking about 100 cigarettes over a lifetime or 0.004 cigarettes per day, which is the level often used to define a never-smoker. The notion that PM_{2.5} *causes* premature death becomes even more implausible when one realizes that a person who smokes 0.2 cigarettes/day has a daily exposure of about 750 µg/m³. If a 10 µg/m³ increase in PM_{2.5} actually *caused* a 0.61 year reduction in life expectancy, equivalent to the claim of Pope, then a 0.2 cigarettes/day smoker would experience about a 45-year reduction in life expectancy, assuming a linear relationship between changes in PM_{2.5} and life expectancy. In actuality, never-smokers and smokers of 0.2 cigarettes/day do not experience any increase in

total death rate or decrease in life expectancy, in spite of a 50-fold greater exposure to PM_{2.5} (10). Furthermore, hundreds of toxicology experiments on both animals and humans have not proven that PM_{2.5} at levels up to 750 µg/m³ *causes* death. Finally, the small relative risks of death and other biases and weaknesses of the PM_{2.5} epidemiologic studies do not meet the standards of causality set by the 2011 Federal Judicial Center Reference Manual on Scientific Evidence (11). The legal standard for causality in epidemiologic studies is a large relative risk (RR ≥ 2.0), not the small relative risk (RR ~ 1.1) typically found in PM_{2.5}-mortality studies.

3) Misrepresentation of PM_{2.5}-death findings has harmed the credibility of epidemiology. The PM_{2.5}-mortality relationship has been contested since 1993 because this small risk could be due to well-known biases, such as, confounding variables and the ecological fallacy. In spite of these biases, several major PM_{2.5} investigators continue to assert that selected positive findings prove that PM_{2.5} *causes* death and they continue to ignore or dismiss null PM_{2.5} results. Enstrom prepared a detailed November 15, 2013 document (5000 words of text with 77 URLs) which describes many misrepresentations and exaggerations (12). In particular, Pope and others have ignored null PM_{2.5} findings in California. Serious concerns about the PM_{2.5}-mortality relationship in California were expressed at a February 26, 2010 Symposium on “Estimating Premature Deaths from Long-term Exposure to PM_{2.5}” by the California Air Resources Board (CARB). Vastly different viewpoints were expressed by scientists like Enstrom and Pope. Although this Symposium could have led to better understanding and cooperation among PM_{2.5} investigators, it did not. For instance, three Symposium attendees (Pope, Jerrett, and Krewski), published extensive findings in their October 28, 2011 CARB report showing that there was an overall null relationship between PM_{2.5} and mortality in California, if one averaged the results from all nine of their models. This null finding agrees exactly with the null findings of Enstrom and others. However, in their subsequent September 1, 2013 *AJRCCM* paper, “Air Pollution and Mortality in California,” they selectively published the positive findings found in one model, but omitted the null findings of the eight other models in their 2011 report.

4) The American Cancer Society actively supports “secret science” PM_{2.5} epidemiology. Since 1995 ACS has repeatedly allowed its 1982 Cancer Prevention Study (CPS II) data to be selectively used for PM_{2.5} epidemiology research. However, ACS has refused to release the CPS II data or allow analysis that addresses the legitimate concerns raised by qualified critics of this “secret science” research. ACS is well aware of the scientific controversy generated by the original 1995 Pope *AJRCCM* paper and subsequent papers that have been used by EPA as a primary justification for its PM_{2.5} regulations. The demand for CPS II data access has increased as PM_{2.5}-related regulations have gotten stricter, more expensive, and more implausible. While ACS refuses any independent access to its CPS II data, because of alleged concerns about subject confidentiality, it has repeatedly allowed Pope and his collaborators to violate a confidentiality pledge made to CPS II subjects. When personal questionnaire data was collected from CPS II subjects upon enrollment in late 1982, ACS informed them with this exact sentence: “We will **never** release information about any particular person and will not release addresses to any agency for any purpose, whatsoever” (13). Both the September 1, 2013 *AJRCCM* paper and the new January 2, 2015 *Circulation Research* paper by Pope include findings based on linking the home address of each study subject to a geographically estimated PM_{2.5} concentration, in violation of the 1982 agreement.

Our evidence that PM_{2.5} does not *cause* premature deaths invalidates the \$1.7 trillion annual benefit that EPA attributes to reductions in PM_{2.5} and supports Cox’s findings that the economic costs of EPA CAA Amendment regulations exceed the resulting health benefits. Because the scientific and economic stakes are high for America, there is an urgent need for transparency and reproducibility in the science and data underlying EPA regulations, as required by the SSRA. The data access requirement in the SSRA is very similar to the one *Science* has for its research papers and to the one recently recommended by the editors of 30 major journals, including *Science* (14). Even an environmental organization that objects to the SSRA, the Union of Concerned Scientists, realizes that “public trust in science increases when we all have access to the same base of evidence” (15).

References

1. U.S.Environmental Protection Agency. “The Benefits and Costs of the Clean Air Act from 1990 to 2020” April 2011
(http://www.epa.gov/air/sect812/feb11/fullreport_rev_a.pdf)
2. L. A. Cox, Jr. Reassessing the human health benefits from cleaner air. *Risk Anal* 2012;32(5):816–829. DOI: 10.1111/j.1539-6924.2011.01698.x. Epub 2011 Nov 2.
3. F. Dominici, M. Greenstone, C. R. Sunstein. Policy Forum: Particulate matter matters. *Science* 2014;344:257-259 (April 18, 2014) DOI: 10.1126/science.1247348
(<http://www.sciencemag.org/content/344/6181/257.full.pdf>)
4. S. Gallagher. AAAS Summary of the Secret Science Reform Act of 2015 (May 7, 2015)
(<http://www.aaas.org/news/summary-secret-science-reform-act-2015>)
5. J. E. Enstrom. Response to Pope “Fine-particulate air pollution and life expectancy in the United States,” May 27, 2009 CARB Comment (http://www.arb.ca.gov/lists/gmove09/1-carb_enstrom_comments_re_pm2.5_and_life_expectancy_052709.pdf)
6. G. Krstić. A reanalysis of fine particulate matter air pollution versus life expectancy in the United States. *J Air Waste Manage Assoc* 2012;62(9):989-991. DOI: 10.1080/10962247.2012.697445
7. S. S. Young, J. Q. Xia. Assessing Geographic Heterogeneity and Variable Importance in an Air Pollution Data Set. *Statistical Analysis and Data Mining* 2013;6:375–386. DOI:10.1002/sam.11202
8. L. A. Cox, Jr., D. A. Popken. Has reducing fine particulate matter and ozone caused reduced mortality rates in the United States? *Annals Epidemiology* 2015; 25:162-173. DOI: 10.1016/j.annepidem.2014.11.006

9. NCHS National Vital Statistics Report (NVSR Volume 64, Number 2) “Deaths: Final Data for 2013.” December 29, 2014
(http://www.cdc.gov/nchs/data/nvsr/nvsr64/nvsr64_02.pdf)
10. J. E. Enstrom, S. S. Young, C. Battig, J. D. Dunn. Tenth International Conference on Climate Change, Panel 8 on Health Effects, June 11, 2015, Washington, DC
(<http://climateconference.heartland.org/sessions/iccc10-panel-8-enstrom-young-battig/>)
and (<http://www.scientificintegrityinstitute.org/JEEICCC061115.pdf>)
11. J. D. Dunn. Climate Conference Human Health Effects Research. JunkScience.com June 9, 2015 (<http://junkscience.com/2015/06/09/72157/>)
12. J. E. Enstrom. Public Comments to EPA on Clean Power Plan, December 1, 2014
(<http://www.scientificintegrityinstitute.org/JEECPP120114.pdf>)
13. J. E. Enstrom. Evidence ACS Violated 1982 CPS II Subject Confidentiality Statements September 2, 2013 (<http://www.scientificintegrityinstitute.org/Enstrom090213.pdf>)
14. M. McNutt. Journals unite for reproducibility. *Science* 2014;346:679 (November 7, 2014)
(<http://www.sciencemag.org/content/346/6210/679.full>)
15. A. A. Rosenberg, L. M. Branscomb, V. Eady, P.C. Frumhoff, G. T. Goldman, M. Halpern, K. Kimmell, Y. Kothari, L. D. Kramer, N. F. Lane, J. J. McCarthy, P. Phartiyal, K. Rest, R. Sims, C. Wexler. Policy Forum: Congress's attacks on science-based rules. *Science* 2015;348:964-966 (May 29, 2015) DOI: 10.1126/science.aab2939
(<http://www.sciencemag.org/content/348/6238/964.full.pdf>)