Comments on US EPA's Integrated Science Assessment for Sulfur Oxides – Health Criteria (First External Review Draft)

Julie E. Goodman, Ph.D., DABT, FACE, ATS
Gradient
January 27, 2016

Thank you for the opportunity to speak on behalf of the American Petroleum Institute on the first draft SO_x Integrated Science Assessment (ISA). Today, I'd like to discuss several areas I think CASAC should consider that would strengthen the ISA. These cover causal determinations and their relevance to the NAAQS, the systemic evaluation of study quality, and proposed modes of action (MoAs) in relation to ambient SO_2 levels.

First, the ISA states, "judgments regarding causality are made by evaluating the evidence over the full range of exposures…defined in this ISA to be relevant to ambient exposure." The ISA defines relevant SO_2 exposures to be those up to 2,000 ppb, and there is little question that short-term, often transitory, respiratory effects can occur at this high concentration. However, is this really the right question? Should the ISA also address whether the current NAAQS is adequate to protect public health? Since the 2008 ISA, there have been no studies that indicate new effects or sensitivities at or below the current NAAQS. CASAC should consider whether the ISA should also incorporate analyses based on the level of the current NAAQS, in addition to exposures up to 2,000 ppb.

Second, the ISA states that the quality of each individual study is evaluated using a set of specific study quality criteria, which are presented in Table A-1. This set of criteria is a welcome addition to the SO_x ISA because it details what must be considered when determining study quality. These criteria most certainly should be used to perform systematic weight-of-evidence analyses. Unfortunately, it appears that these criteria are used inconsistently throughout the ISA.

It would be helpful for CASAC to discuss how the ISA considers these criteria when evaluating individual studies and how these criteria could be applied in a consistent and systematic manner. I have included a table Gradient put together to show how one could evaluate studies systematically and transparently, using short-term SO_2 exposure and hospital admissions (HAs) and emergency department (ED) visits for asthma as an example. We present each of the 17 studies according to various study quality characteristics listed in Table A-1, and highlight in green the characteristics that the draft ISA considers to be are indicative of a higher quality study and pink for those that are considered limitations (*i.e.*, not meeting the explicitly stated study quality criteria). This is not the only way these characteristics can be laid out, but something like this should be considered to ensure consistent evaluations across studies.

Third, as the ISA indicates, it is critical to evaluate MoAs when assessing causality. The question at issue is whether the proposed MoAs are complete and indicate that effects are biologically plausible. One challenge is to identify and discuss whether there are critical data gaps that impede a clear determination of plausibility. This includes an evaluation of the impact of dosimetry, pharmacokinetics, and pharmacodynamics on the proposed key events, as well as a discussion of which experimental conditions

are relevant to humans. For example, endogenous concentrations of SO_2 products far exceed the minimal amount generated from SO_2 inhalation, yet systemic effects from SO_2 are considered by the ISA, and several MoAs are considered, but not fully evaluated. CASAC is in a unique position to work through these issues and objectively decide whether all the mechanistic evidence (including that regarding homeostatic mechanisms) has been adequately integrated, and whether the ISA's causal determinations are supported by this evidence.

Finally, I ask CASAC to consider whether evidence that was determined to be inadequate in 2008 is now truly suggestive of causation for several health effects. The ISA acknowledges that the studies published since 2008 have many of the same uncertainties as those already considered. As I have discussed at prior CASAC meetings, the language used to describe the ISA's causal determinations can easily be misinterpreted. Even though definitions are provided, research demonstrates a descriptive word like "suggestive" can be interpreted to mean it is more likely than not that a causal effect occurs following SO_2 exposure. It is clear that this is not the conclusion the ISA intends to convey. As more focus is placed on communicating scientific results to the public, we should consider revising these categories. Some options to consider are consolidating this category with "inadequate" (as the ISA does for at-risk factors), providing qualitative probability ranges, or changing the words used to define categories.

I will expand on these points in my written comments, and welcome any questions you may have and look forward to your discussions at this meeting. Thank you.

Table A-1 Scientific considerations for evaluating the strength of inference from studies on the health effects of sulfur oxides.

Study Design
Controlled Human Exposure:
Studies should clearly describe the primary and any secondary objectives of the study, or specific hypotheses being tested. Study subjects should be randomly exposed without knowledge of the exposure condition. Preference is given to balanced crossover (repeated measures) or parallel design studies that include control exposures (e.g., to clean filtered air). In crossover studies, a sufficient and specified time between exposure days should be employed to avoid carry over effects from prior exposure days. In parallel design studies, all arms should be matched for individual characteristics such as age, sex, race, anthropometric properties, and health status. In studies evaluating effects of disease, appropriately matched healthy controls are desired for interpretative purposes.
Animal Toxicology:
Studies should clearly describe the primary and any secondary objectives of the study, or specific hypotheses being tested. Studies should include appropriately matched control exposures (e.g., to clean filtered air, time matched). Studies should use methods to limit differences in baseline characteristics of control and exposure groups. Studies should randomize assignment to exposure groups and where possible conceal allocation from research personnel. Groups should be subjected to identical experimental procedures and conditions; animal care including housing, husbandry, etc. should be identical between groups. Blinding of research personnel to study group may not be possible due to animal welfare and experimental considerations; however, differences in the monitoring or handling of animals in all groups by research personnel should be minimized.
Epidemiology:
Inference is stronger for studies that clearly describe the primary and any secondary aims of the study, or specific hypotheses being tested. For short-term exposure, time-series, case crossover, and panel studies are emphasized over cross-sectional studies because they examine temporal correlations and are less prone to confounding by factors that differ between individuals (e.g., SES, age). Studies with large sample sizes and conducted over multiple years are considered to produce more reliable results. If other quality parameters are equal, multicity studies carry more weight than single-city studies because they tend to have larger sample sizes and lower potential for publication bias. For long-term exposure, inference is considered to be stronger for prospective cohort studies and case-control studies nested within a cohort (e.g., for rare diseases) than cross-sectional, other case-control, or ecologic studies. Cohort studies can better inform the temporality of exposure and effect. Other designs can have uncertainty related to the appropriateness of the control group or validity of inference about individuals from group-level data. Study design limitations can bias health effect associations in either direction.

Table A-1 (Continued): Scientific considerations for evaluating the strength of inference from studies on the health effects of sulfur oxides.

Study Population/Test Model
Controlled Human Exposure:
In general, the subjects recruited into study groups should be similarly matched for age, sex, race, anthropometric properties, and health status. In studies evaluating effects of specific subject characteristics (e.g., disease, genetic polymorphism, etc.), appropriately matched healthy controls are preferred. Relevant characteristics and health status should be reported for each experimental group. Criteria for including and excluding subjects should be clearly indicated. For the examination of populations with an underlying health condition (e.g., asthma), independent, clinical assessment of the health condition is ideal, but self report of physician diagnosis generally is considered to be reliable for respiratory and cardiovascular disease outcomes.[a] The loss or withdrawal of recruited subjects during the course of a study should be reported. Specific rationale for excluding subject(s) from any portion of a protocol should be explained.
Animal Toxicology:
Ideally, studies should report species, strain, substrain, genetic background, age, sex, and weight. Unless data indicate otherwise, all animal species and strains are considered appropriate for evaluating effects of SO_2 exposure. It is preferred that the authors test for effects in both sexes and multiple lifestages, and report the result for each group separately. All animals used in a study should be accounted for, and rationale for exclusion of animals or data should be specified.
Epidemiology:
Confidence in results is greater in studies that recruit the study population from the target population and examine a study population that is representative of the target population. Studies with high participation and low drop-out over time that is not dependent on exposure or health status are considered to have low potential for selection bias. Clear indication of criteria for including and excluding subjects can facilitate assessment of selection bias. For populations with an underlying health condition, independent, clinical assessment of the health condition is valuable, but self report of physician diagnosis generally is considered to be reliable for respiratory and cardiovascular outcomes.a Comparisons of groups with and without an underlying health condition are more informative if groups are from the same source population. Selection bias can influence results in either direction or may not affect the validity of results but rather reduce the generalizability of findings to the target population.
Pollutant
Controlled Human Exposure:
The focus is on studies testing SO_2 exposure.
Animal Toxicology:
The focus is on studies testing SO_2 exposure.
Epidemiology:
The focus is on studies testing SO_2 exposure.

Table A-1 (Continued): Scientific considerations for evaluating the strength of inference from studies on the health effects of sulfur oxides.

Exposure Assessment or Assignment
Controlled Human Exposure:
For this assessment, the focus will be on studies that utilize SO_2 concentrations less than or equal to 2 ppm (Section 1.2). Studies that utilize higher exposure concentrations may provide information relevant to mode of action, dosimetry, inter-species variation, or at-risk human populations. Controlled human exposure studies considering short-term peak exposures, defined here as exposures from 5–10 minutes, to 0.2–0.6 ppm SO_2, were emphasized (Section 1.2).
Animal Toxicology:
For this assessment, the focus will be on studies that utilize SO_2 concentrations less than or equal to 2,000 ppb (Section 1.2). Studies that utilize higher exposure concentrations may provide information relevant to mode of action, dosimetry, inter-species variation, or at-risk human populations. Studies should characterize pollutant concentration, temperature, and relative humidity and/or have measures in place to adequately control the exposure conditions. The focus is on inhalation exposure. Noninhalation exposure experiments may provide information relevant to mode of action. In vitro studies may be included if they provide mechanistic insight or examine similar effects as in vivo, but are generally not included. All studies should include exposure control groups (e.g., clean filtered air).
Epidemiology:
Of primary relevance are relationships of health effects with the ambient component of exposure to SO_2. However, information about ambient exposure rarely is available for individual subjects; most often, inference is based on ambient concentrations. Studies that compare exposure assessment methods are considered to be particularly informative. Inference is stronger when the duration or lag of the exposure metric corresponds with the time course for physiological changes in the outcome (e.g., up to a few days for symptoms) or latency of disease (e.g., several years for cancer). Given the spatial heterogeneity in ambient SO_2 and potentially variable relationships between personal exposures and ambient concentrations (Sections 3.3.2 and 3.3.3.1), validated methods that capture the extent of variability for the particular study design (temporal vs. spatial contrasts) and location carry greater weight. Central site measurements, whether averaged across multiple monitors or assigned from the nearest or single available monitor, have well-recognized limitations in capturing spatial variation in air pollutants. Monitors impacted by large SO_2 sources are particularly subject to concentration fluctuations due to changes in emission rates and meteorological conditions and may not fully represent population exposure. Results based on central site measurements can be informative if correlated with personal exposures, closely located to study subjects, highly correlated across monitors within a location, used in locations with well-distributed sources, or combined with time-activity information. In studies of short-term exposure, temporal variability of the exposure metric is of primary interest. Metrics that may capture variation in ambient sulfur oxides and strengthen inference include concentrations in subjects' microenvironments and individual-level outdoor concentrations combined with time-activity data. Atmospheric models may be used for exposure assessment in place of or to supplement SO_2 measurements in epidemiologic analyses. Dispersion models (e.g., AERMOD) can provide valuable information on fine-scale temporal and spatial variations (within tens of km) of SO_2 concentrations, which is particularly important for assessing exposure near large stationary sources. Alternatively, grid-scale models (e.g., CMAQ) that represent SO_2 exposure over relatively large spatial scales (e.g., typically greater than 4 × 4 km grid size) often do not provide enough spatial resolution to capture acute SO_2 peaks that influence short-term health outcomes. Uncertainty in exposure predictions from these models is largely influenced by model formulations and the quality of model input data pertaining to emissions or meteorology, which tends to vary on a study-by-study basis. For long-term exposures, models that capture within-community spatial variation in individual exposure may be given more weight for spatially variable ambient SO_2. Exposure measurement error often attenuates health effect estimates or decreases the precision of the association (i.e., wider 95% CIs), particularly associations based on temporal variation in short-term exposure (Section 3.3.5.1). However, exposure measurement error can bias estimates away from the null, particularly for long-term exposures.

Table A-1 (Continued): Scientific considerations for evaluating the strength of inference from studies on the health effects of sulfur oxides.

Outcome Assessment/Evaluation
Controlled Human Exposure:
Endpoints should be assessed in the same manner for control and exposure groups (e.g., time after exposure, methods, endpoint evaluator) using valid, reliable methods. Blinding of endpoint evaluators is ideal, especially for qualitative endpoints (e.g., histopathology). For each experiment and each experimental group, including controls, precise details of all procedures carried out should be provided including how, when, and where. Time of the endpoint evaluations is a key consideration that will vary depending on the endpoint evaluated. Endpoints should be assessed at time points that are appropriate for the research questions.
Animal Toxicology:
Endpoints should be assessed in the same manner for control and exposure groups (e.g., time after exposure, methods, endpoint evaluator) using valid, reliable methods. Blinding of endpoint evaluators is ideal, especially for qualitative endpoints (e.g., histopathology). For each experiment and each experimental group, including controls, precise details of all procedures carried out should be provided including how, when, and where. Time of the endpoint evaluations is a key consideration that will vary depending on the endpoint evaluated. Endpoints should be assessed at time points that are appropriate for the research questions.
Epidemiology:
Inference is stronger when outcomes are assessed or reported without knowledge of exposure status. Knowledge of exposure status could produce artifactual associations. Confidence is greater when outcomes assessed by interview, self report, clinical examination, or analysis of biological indicators are defined by consistent criteria and collected by validated, reliable methods. Independent, clinical assessment is valuable for outcomes such as lung function or incidence of disease, but report of physician diagnosis has shown good reliability.[a] Outcomes assessed at time intervals that correspond with the time course for physiological changes (e.g., up to a few days for symptoms) are emphasized. When health effects of long-term exposure are assessed by acute events such as symptoms or hospital admissions, inference is strengthened when results are adjusted for short-term exposure. Validated questionnaires for subjective outcomes such as symptoms are regarded to be reliable,[b] particularly when collected frequently and not subject to long recall. For biological samples, the stability of the compound of interest and the sensitivity and precision of the analytical method is considered. If not based on knowledge of exposure status, errors in outcome assessment tend to bias results toward the null.
Potential Copollutant Confounding
Controlled Human Exposure:
Exposure should be well characterized to evaluate independent effects of SO_2.
Animal Toxicology:
Exposure should be well characterized to evaluate independent effects of SO_2.

Table A-1 (Continued): Scientific considerations for evaluating the strength of inference from studies on the health effects of sulfur oxides.

Epidemiology:
Not accounting for copollutant confounding can produce artifactual associations; thus, studies that examine copollutant confounding carry greater weight. The predominant method is copollutant modeling, which is especially informative when measurement error is comparable for copollutants and correlations are not high. Interaction and joint effect models are examined to a lesser extent. Evaluating correlations between SO_2 and copollutants and comparing health associations between SO_2 and copollutants in single-pollutant models can add to the analysis of potential copollutant confounding, particularly when exposure measurement error is comparable among pollutants. Studies that examine SO_2 only in single-pollutant models provide minimal information on the potential for copollutant confounding. Copollutant confounding is evaluated based on the extent of observed correlations and relationships with health effects. Highly variable correlations have been observed between SO_2 and other criteria pollutants at collocated monitors (Section 2.5.5), ranging from negative to strong correlations, making evaluation of copollutant confounding necessary on a study-specific, rather than a general, basis.
Other Potential Confounding Factors[e]
Controlled Human Exposure:
Preference is given to studies utilizing experimental and control groups that are matched for individual level characteristics (e.g., body weight, smoking history, age) and time-varying factors (e.g., seasonal and diurnal patterns).
Animal Toxicology:
Preference is given to studies utilizing experimental and control groups that are matched for individual level characteristics (e.g., body weight, litter size, food and water consumption) and time-varying factors (e.g., seasonal and diurnal patterns).
Epidemiology:
Factors are considered to be potential confounders if demonstrated in the scientific literature to be related to health effects and correlated with SO_2. Not accounting for confounders can produce artifactual associations; thus, studies that statistically adjust for multiple factors or control for them in the study design are emphasized. Less weight is placed on studies that adjust for factors that mediate the relationship between SO_2 and health effects, which can bias results toward the null. In the absence of information linking health risk factors to SO_2, a factor may be evaluated as a potential effect measure modifier, but uncertainty is noted as to its role as a confounder. Confounders vary according to study design, exposure duration, and health effect and may include, but are not limited to, the following: For time-series and panel studies of short-term exposure: • Respiratory effects—meteorology, day of week, season, medication use, allergen exposure (potential effect modifier) • Cardiovascular effects—meteorology, day of week, season, medication use • Total mortality—meteorology, day of week, season, long-term temporal trends For studies of long-term exposure: • Respiratory effects—socioeconomic status, race, age, medication use, smoking, stress • Cardiovascular, reproductive, and development effects—socioeconomic status, race, age, medication use, smoking, stress, noise • Total mortality—socioeconomic status, race, age, medication use, smoking, comorbid health conditions • Cancer—socioeconomic status, race, age, occupational exposure

Table A-1 (Continued): Scientific considerations for evaluating the strength of inference from studies on the health effects of sulfur oxides.

Statistical Methodology
Controlled Human Exposure:
Statistical methods should be clearly described and appropriate for the study design and research question (e.g., correction for multiple comparisons). Generally, statistical significance is used to evaluate the findings of controlled human exposure studies. Detection of statistical significance is influenced by a variety of factors including, but not limited to, the size of the study, exposure and outcome measurement error, and statistical model specifications. Sample size is not a criterion for exclusion; ideally, the sample size should provide adequate power to detect hypothesized effects (e.g., sample sizes less than three are considered less informative). Because statistical tests have limitations, consideration is given to both trends in data and reproducibility of results.
Animal Toxicology:
Statistical methods should be clearly described and appropriate for the study design and research question (e.g., correction for multiple comparisons). Generally, statistical significance is used to evaluate the findings of animal toxicology studies. Detection of statistical significance is influenced by a variety of factors including, but not limited to, the size of the study, exposure and outcome measurement error, and statistical model specifications. Sample size is not a criterion for exclusion; ideally, the sample size should provide adequate power to detect hypothesized effects (e.g., sample sizes less than three are considered less informative). Because statistical tests have limitations, consideration is given to both trends in data and reproducibility of results.
Epidemiology:
Multivariable regression models that include potential confounding factors are emphasized. However, multipollutant models (more than two pollutants) are considered to produce too much uncertainty due to copollutant collinearity to be informative. Models with interaction terms aid in the evaluation of potential confounding as well as effect modification. Sensitivity analyses with alternate specifications for potential confounding inform the stability of findings and aid in judgments of the strength of inference of results. In the case of multiple comparisons, consistency in the pattern of association can increase confidence that associations were not found by chance alone. Statistical methods that are appropriate for the power of the study carry greater weight. For example, categorical analyses with small sample sizes can be prone to bias results toward or away from the null. Statistical tests such as t-tests and Chi-squared tests are not considered sensitive enough for adequate inferences regarding pollutant-health effect associations. For all methods, the effect estimate and precision of the estimate (i.e., width of 95% CI) are important considerations rather than statistical significance.

AERMOD = American Meteorological Society/U.S. EPA Regulatory Model; CI = confidence interval; CMAQ = Community Multiscale Air Quality; ppb = parts per billion; ppm = parts per million; SES = socioeconomic status; SO_2 = sulfur dioxide.

[a]Toren et al. (1993); (Murgia et al. (2014); Weakley et al. (2013); Yang et al. (2011); Heckbert et al. (2004); Barr et al. (2002); Muhajarine et al. (1997)).

[b]Burney et al. (1989).

[c]Many factors evaluated as potential confounders can be effect measure modifiers (e.g., season, comorbid health condition) or mediators of health effects related to SO_2 (comorbid health condition).

Asthma HA/ED Visit Study Quality Characteristics

Citation	Study Design			Exposure Assessment			Outcome Assessment		Confounding by Co-pollutants			Other Confounders				Statistical Methods
	Design	Single *vs*. Multi-city	Size/Duration[1]	Central Site Monitoring	Spatial Variability Assessed	Comparison of Exposure Assessment Methods	Type of Outcome	Exclusion of Children < 2 Years Old	Co-pollutants Assessed	Correlations Reported	Relative Measurement Error in Co-pollutants Discussed	Meteorology	Day of Week	Season	Allergens	Sensitivity Analysis: Alternate Model Specification
Strickland *et al.* (2010)	Time series and case cross-over	Single	91,386 ED visits/ 12 years	Yes	No	No	ED visits	Yes	No[2]	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Villeneuve *et al.* (2007)	Case cross-over	Single	57,912 ED visits/ 10 years	Yes	No	No	ED visits	Yes	No[2]	Yes	Yes	Yes	Yes	Yes	Yes	No
Jalaludin *et al.* (2008)	Case cross-over	Single	5 years	Yes	No	No	ED visits	No[3]	CO, $PM_{2.5}$, PM_{10}, NO_2, O_3	Yes	No	Yes	Yes	Yes	No	No
Ito *et al.* (2007)	Time series	Single	4 years	Yes	Yes	No	ED visits	No	CO, $PM_{2.5}$, NO_2, O_3,	No	Yes	Yes	Yes	Yes	No	Yes
Stieb *et al.* (2009)	Time series	Multi-city	4-10 years[4]	Yes	No	No	ED visits	No	No[2]	Yes	No	Yes	Yes	Yes	No	Yes
Samoli *et al.* (2011)	Time series	Single	4 years	Yes	Yes	No	HAs	No	PM_{10}, NO_2, O_3	Yes	No	Yes	Yes	Yes	No[5]	Yes
Peel *et al* . (2005)	Time series	Single	8 years	Yes	Yes	Yes[6]	ED visits	No[7]	No[8]	Yes	No	Yes	Yes	Yes	Yes	Yes
ATSDR and NYSDOH (2006)	Time series	Single	2 years	Yes	Yes	No	ED visits	No[9]	$PM_{2.5}$, NO_2, O_3	Yes	No	Yes	Yes	Yes	No[10]	Yes
Son *et al.* (2013)	Time series	Multi-city	6 years	Yes	No	No	HAs	No	No[2]	Yes	No	Yes	Yes	Yes	No	Yes
Orazzo *et al.* (2009)	Case cross-over	Multi-city	53,272 ED visits/ 7 years	Yes	Yes	No	ED visits	No	No[2]	No	No	Yes	Yes	Yes	No[10]	Yes
Li *et al.* (2011)	Time series and case cross-over	Single	12,933 asthma events/ 3 years	Yes	No	No	ED visits and HAs	Yes	No[2]	Yes	No	Yes	Yes	Yes	No	Yes
Winquist *et al.* (2014)	Time series	Single	6 years	Yes	No	No	ED visits	Yes	NO_2, O_3, SO_4^{2-}, CO, $PM_{2.5}$	Yes	Yes	Yes	Yes	Yes	No	Yes
Jaffe *et al.* (2003)	Time series	Multi-city	6 summers	Yes	No[11]	No	ED visits	Yes	No[2]	Yes	No	Yes	Yes	Yes	No	No
Lin *et al.* (2004)	Case-control	Single	2,629 asthma HAs/2.5 years	Yes	No	No	HAs	No	No	NA	NA	Yes	Yes	Yes	No	No
Sheppard (2003) and Sheppard *et al.* (1999)[12]	Time series	Single	8 years	Yes	No	No	HAs	No	No[2]	Yes	No	Yes	Yes	Yes	No	Yes
Wilson *et al.* (2005)	Time series	Multi-city	3-5 years[8]	Yes	No	No	ED visits	No	No[2]	No	No	Yes	Yes	Yes	No	Yes
Smargiassi *et al.* (2009)	Case cross-over	Single	9 years	No[13]	Yes	Yes	ED visits and HAs	Yes	O_3, NO_2, SO_2, $PM_{2.5}$[14]	No	No	Yes	Yes	Yes	No	No

Notes: Green shading indicates characteristics that the ISA states are indicative of a higher quality study. Pink indicates characteristics that are considered limitations (*i.e.* , not meeting the study quality criteria).

CO = Carbon Monoxide; ED = Emergency Department; HA = Hospital Admission; ISA = Integrated Science Assessment; NO_2 = Nitrogen Dioxide; O_3 = Ozone; PM = Particulate Matter; PM_{10} = particulate matter less than 10 µm in diameter; $PM_{2.5}$ = particulate matter less than 2.5 µm in diameter; SO_4^{2-} = sulfate; SO_2 = Sulfur Dioxode; SO_x = Oxides of Sulfur.

(1) In Table A-1, US EPA did not indicate what sample size and duration are required for a study to be considered "large" and, therefore, more reliable.
(2) Several co-pollutants were measured and examined in single-pollutant models. SO_2 was not included in any two-pollutant models.
(3) < 1-year-old subjects excluded.
(4) Duration varied by city.
(5) Desert dust, which includes bio-allergens.
(6) Compared monitoring systems.
(7) 0- to 1-year-old subjects analyzed separately.
(8) The authors conducted several multi-pollutant analyses, but did not present any results for models that included SO_x.
(9) Included additional diagnostic criteria for children < 1 year old to mitigate outcome misclassification.
(10) Aeroallergens measured but not included in statistical models.
(11) Spatial variability of other pollutants (*e.g.,* PM) was assessed, but spatial variability of SO_2 was not.
(12) Sheppard (2003) is a reanalysis of data from Sheppard *et al. (* 1999).
(13) In addition to two central-site monitors, resident-level exposure estimates were calculated using the AERMOD dispersion model.
(14) Authors controlled for regional background concentrations.