Risk and Exposure Assessment for the Review of the Primary National Ambient Air Quality Standard for Sulfur Oxides

EPA-452/R-18-003
May 2018

Risk and Exposure Assessment for the Review of the Primary National Ambient Air Quality Standard for Sulfur Oxides

U.S. Environmental Protection Agency
Office of Air Quality Planning and Standards
Health and Environmental Impacts Division
Research Triangle Park, NC

DISCLAIMER

This document has been prepared by staff in the Health and Environmental Impacts Division, Office of Air Quality Planning and Standards, U.S. Environmental Protection Agency (EPA). Any findings and conclusions are those of the authors and do not necessarily reflect the views of the Agency. Mention of trade names or commercial products is not intended to constitute endorsement or recommendation for use. Questions or comments related to this document should be addressed to Dr. Stephen Graham, U.S. Environmental Protection Agency, Office of Air Quality Planning and Standards, C539-07, Research Triangle Park, North Carolina 27711 (email: graham.stephen@epa.gov) and Dr. Nicole Hagan, U.S. Environmental Protection Agency, Office of Air Quality Planning and Standards, C504-06, Research Triangle Park, North Carolina 27711 (email: hagan.nicole@epa.gov).

TABLE OF CONTENTS

LIST OF APPENDICES

LIST OF FIGURES

LIST OF TABLES

LIST OF ACRONYMS AND ABBREVIATIONS

A/C	air conditioner
ACS	American Community Survey
AER	air exchange rate
AHR	airway hyperresponsiveness
AHS	American Housing Survey
APEX	Air Pollutants Exposure model
AQS	Air Quality System
ASOS	Automated Surface Observing Stations
BASE	Building Assessment Survey and Evaluation
BSA	body surface area
CAA	Clean Air Act
CASAC	Clean Air Scientific Advisory Committee
CHAD	Consolidated Human Activity Database
DV	design value
EGU	Electricity generating unit
EPA	Environmental Protection Agency
E-R	exposure-response
EVR	equivalent ventilation rate
FEV_1	forced expiratory volume in one minute
IRP	Integrated Review Plan
ISA	Integrated Science Assessment
ISH	Integrated Surface Hourly
km	kilometer
lat	latitude
lon	longitude
m	meter
MCC	Markov-chain clustering
ME	microenvironment
MER	mixed-effects regression
MLR	multiple linear regression
MRLC	Multi-Resolution Land Characteristics
MSA	Metropolitan Statistical Area
NAAQS	National Ambient Air Quality Standard
NCEI	National Centers for Environmental Information
NED	National Elevation Data

NEI	National Emissions Inventory
NHIS	National Health Interview Survey
NLCD	National Land Cover Dataset
NO_2	nitrogen dioxide
NWS	National Weather Service
O_3	ozone
OAQPS	Office of Air Quality Planning and Standards
ppb	parts per billion
PA	Policy Assessment
PM	particulate matter
PMR	peak-to-mean ratio
PSD	Prevention of Significant Deterioration
REA	Risk and Exposure Assessment
RMR	resting metabolic rate
SIP	State Implementation Plan
SO_X	oxides of sulfur
sRaw	specific airway resistance
$\dot{V}_E$	activity-specific ventilation rate
WHO	World Health Organization

1 INTRODUCTION

This document, *Risk and Exposure Assessment for the Review of the Primary National Ambient Air Quality Standard for Sulfur Oxides* (hereafter referred to as *REA*), describes the quantitative human exposure and risk characterization conducted to inform the U.S. Environmental Protection Agency's (EPA's) current review of the primary (health-based)[1] national ambient air quality standard (NAAQS) for sulfur oxides (SO_X). This document presents the methods, key results, observations, and related uncertainties associated with the quantitative analyses performed. The REA draws upon the Integrated Science Assessment (ISA; U.S. EPA 2017a) and reflects consideration of the Clean Air Scientific Advisory Committee's (CASAC) advice and public comments on the draft REA.

In this review, as in each NAAQS review, the policy implications of the REA results are considered in the policy assessment prepared separately for the review. The policy assessment presents analyses and staff conclusions regarding the policy implications of the key scientific and technical information that informs the review. The policy assessment is intended to "bridge the gap" between the relevant scientific evidence and technical information and the judgments required of the Administrator in his consideration of the adequacy of the current standards. The policy assessment for this review of the primary NAAQS for SO_X is titled, *Policy Assessment for the Review of the Primary National Ambient Air Quality Standard for Sulfur Oxides* (PA; U.S. EPA, 2018).

The remainder of this chapter summarizes the legislative requirements (section 1.1), provides an overview of the history of the primary NAAQS for SO_X (section 1.2), and describes aspects of the REA that have been updated since the 2009 REA, and revisions made from the draft to the final REA in consideration of CASAC recommendations and public comments (section 1.3). Following Chapter 1, the REA presents an overview of the assessment approach (Chapter 2), describes the study areas and air quality modeling (Chapter 3), describes the exposure modeling and risk characterization (Chapter 4), presents the exposure and risk estimates (Chapter 5), and describes the analysis of variability and characterization of uncertainty (Chapter 6).

---

[1] The EPA is separately reviewing the welfare effects associated with sulfur oxides and the public welfare protection provided by the secondary SO_2 standard, in conjunction with a review of the secondary standards for nitrogen oxides and particulate matter with respect to their protection of the public welfare from adverse effects related to ecological effects (U.S. EPA, 2017b).

1.1 BACKGROUND

Sections 108 and 109 of the Clean Air Act (CAA) govern the establishment and periodic review of the NAAQS. Section 108 [42 U.S.C. 7408] directs the Administrator to identify and list certain air pollutants and then to issue air quality criteria for those pollutants. The Administrator is to list those air pollutants "emissions of which, in his judgment, cause or contribute to air pollution which may reasonably be anticipated to endanger public health or welfare," "the presence of which in the ambient air results from numerous or diverse mobile or stationary sources;" and "for which…[the Administrator] plans to issue air quality criteria…." CAA section 108(a)(1). The NAAQS are established for the pollutants listed. The CAA requires that NAAQS are to be based on air quality criteria, which are intended to "accurately reflect the latest scientific knowledge useful in indicating the kind and extent of all identifiable effects on public health or welfare which may be expected from the presence of [the] pollutant in the ambient air…" CAA section 108(a)(2). Under CAA section 109 [42 U.S.C. 7409], the EPA Administrator is to propose, promulgate, and periodically review, at five-year intervals, "primary" (health-based) and "secondary" (welfare-based)[2] NAAQS for such pollutants for which air quality criteria are issued.[3] Based on periodic reviews of the air quality criteria and standards, the Administrator is to make revisions in the criteria and standards, and promulgate any new standards, as may be appropriate. The CAA also requires that an independent scientific review committee review the air quality criteria and standards and recommend to the Administrator any new standards and revisions of existing air quality criteria and standards as may be appropriate, a function now performed by the CASAC.

The current primary NAAQS for SO_X is a 1-hour standard set at a level of 75 parts per billion (ppb), based on the 3-year average of the annual 99th percentile of 1-hour daily maximum SO_2 concentrations. This standard was set in the last review of the primary NAAQS for SO_X, which was completed in 2010 (75 FR 35520, June 22, 2010). In comparison to the standards existing at that time, establishment of the 1-hour standard was determined to provide increased protection for people with asthma and other at-risk populations against an array of respiratory

---

[2] Section 302(h) of the CAA provides that all language referring to effects on welfare includes but is not limited to, "…effects on soils, water, crops, vegetation, man-made materials, animals, wildlife, weather, visibility, and climate, damage to and deterioration of property, and hazards to transportation, as well as effects on economic values and on personal comfort and well-being…."

[3] Section 109(b)(1) [42 U.S.C. 7409] of the CAA defines a primary standard as one "the attainment and maintenance of which in the judgment of the Administrator, based on such criteria and allowing an adequate margin of safety, are requisite to protect the public health." Section 109(b)(2) of the CAA directs that a secondary standard is to "specify a level of air quality the attainment and maintenance of which, in the judgment of the Administrator, based on such criteria, is requisite to protect the public welfare from any known or anticipated adverse effects associated with the presence of [the] pollutant in the ambient air."

effects related to short-term exposures (as short as 5 minutes) and to maintain longer-term concentrations below those specified by the then-existing standards (75 FR 35550, June 22, 2010).[4]

The EPA initiated the current review of the primary NAAQS for SO_X in May 2013, with a call for information from the public (78 FR 27387, May 10, 2013). The EPA held a workshop on June 12-13, 2013 to discuss policy-relevant scientific and technical information to inform the EPA's planning for the review. Following the workshop, the EPA developed the plan for the review, which is described in the *Integrated Review Plan for the Primary National Ambient Air Quality Standard for Sulfur Dioxide* (U.S. EPA, 2014; hereafter referred to as the IRP). The IRP includes policy-relevant questions for the review, the process and schedule for conducting the review, and descriptions of the purpose, contents and approach for developing the key documents for the review.

The key documents in the review include an Integrated Science Assessment (ISA), a REA (as warranted), and a PA. In general terms, the ISA is to provide a critical assessment of the latest available scientific information upon which the NAAQS are to be based, and the PA is to evaluate the policy implications of the information contained in the ISA and of any policy-relevant quantitative analyses, such as a quantitative REA performed for the current review or, as applicable, for past reviews. Based on that evaluation, the PA presents staff conclusions regarding policy options for the Administrator to consider in reaching decisions on the NAAQS.[5]

The EPA has developed this REA describing the quantitative risk and exposure assessment being conducted by the Agency to support this review of the primary SO_X standard. This document is intended to be a concise presentation of the methods, key results, observations, and related uncertainties associated with the analyses performed. The REA builds upon the health effects evidence presented in the ISA, as well as CASAC advice and public comments on the REA planning document (*Review of the Primary National Ambient Air Quality Standard for Sulfur Oxides: Risk and Exposure Assessment Planning Document,* REA Planning Document, U.S. EPA, 2017c) following a consultation with the CASAC at a public meeting in March 2017 (82 FR 11449). In consideration of CASAC comments at that consultation and public comments, the EPA developed the draft REA (U.S. EPA, 2017d) and the draft PA (U.S. EPA 2017e), which

[4] In the 2010 decision to establish a new 1-hour standard, the EPA revoked the then-existing 24-hour and annual primary standards.

[5] The basic elements of a standard include the indicator, averaging time, form, and level. The indicator defines the pollutant to be measured in the ambient air for the purpose of determining compliance with the standard. The averaging time defines the time period over which air quality measurements are to be obtained and averaged or cumulated. The form of a standard defines the air quality statistic that is to be compared to the level of the standard in determining whether an area attains the standard. The level of a standard defines the air quality concentration used (i.e., an ambient air concentration of the indicator pollutant).

were released on August 4, 2017 (82 FR 43756, September 19, 2017). The draft REA and draft PA were reviewed by the CASAC on September 18-19, 2017 (82 FR 37213, August 9, 2017). Following a CASAC teleconference on April 20, 2018 (83 FR 14638, April 5, 2018), the CASAC's recommendations, based on its review of the draft REA and draft PA, were provided in a letter to the EPA Administrator (Cox and Diez Roux, 2018a,b). The EPA staff considered these recommendations, as well as public comments provided on the draft REA and draft PA, when developing this REA.

The ISA and REA informed the development of the PA and will inform the subsequent rulemaking steps that will lead to final decisions on the primary NAAQS for SO_X. The PA document includes staff analysis of the scientific basis for policy options for consideration by the Administrator prior to rulemaking. The PA integrates and interprets information from the ISA and the REA to frame policy options for consideration by the Administrator. The PA is intended to help "bridge the gap" between the Agency's scientific and technical assessments, presented in the ISA and REA and the judgments required of the Administrator in determining whether it is appropriate to retain or revise the standards. The PA is also intended to facilitate the CASAC's advice to the Administrator on the adequacy of existing standards, and any new standards or revisions to existing standards as may be appropriate. Concurrent with the release of this REA, the PA (U.S. EPA, 2018) is also being released.

The schedule for completion of this review is governed by a court order, which resulted from the entry of consent decree resolving a lawsuit that was filed in July 2016 and that concerned, in relevant part, the timing of completion of this review. *Center for Biological Diversity et al. v. McCarthy* (No. 4:16-cv-07396-VC, N.D. Cal.). The order specifies that the Administrator shall sign a notice setting forth his proposed decision concerning the review of the primary NAAQS for SO_X no later than May 25, 2018; and sign a notice setting forth his final decision concerning the review of the primary NAAQS for SO_X no later than January 28, 2019.

1.2 PREVIOUS REVIEWS AND ASSESSMENTS

Reviews of the primary NAAQS for SO_X completed in 1996 and 2010 included analyses of potential exposure to SO_2 in ambient air (61 FR 25566, May 22, 1996; 75 FR 35520, June 22, 2010). These analyses pertained to the then-existing 24-hour and annual standards, but primarily focused on whether additional protection was necessary to protect at-risk populations (people with asthma) against short-term (e.g., 5-minute) peak exposures while at elevated breathing rates (e.g., while exercising). The analyses that informed the review completed in 1996 focused on potential exposures to 5-minute concentrations at or above 600 ppb for several air quality scenarios (61 FR 2556, May 22, 1996). The 2010 review analyses estimated the number of individuals and percent of the modeled at-risk population that would be expected to experience

5-minute exposures above several concentrations of potential concern extending down to 100 ppb ("benchmark concentrations" based on findings from controlled human exposure studies) and also the number of individuals and percent of the population expected to experience a doubling or greater increase in specific airway resistance (sRaw) or a reduction in forced expiratory volume in one second (FEV_1) of at least 15% (U.S. EPA, 2009 [hereafter referred to as the 2009 REA]). As summarized in more detail in the PA, the analyses in the 2009 REA informed the 2010 decision to establish a new 1-hour standard to protect at-risk populations from short-term (e.g., 5-minute) peak exposures (75 FR 35520, June 22, 2010).

The multiple quantitative analyses that informed the 1996 review decision are described in the 1986 *Addendum to the 1982 OAQPS Staff Paper* (U.S. EPA, 1986), the 1994 *Supplement to the 1986 OAQPS Staff Paper Addendum* (U.S. EPA, 1994) and the final decision notice (61 FR 25566, May 22, 1996). A key aspect of the design for those analyses was the focus on 5-minute concentrations at or above 600 ppb, an exposure level that the Agency judged could pose an immediate significant health risk for a substantial portion of asthmatics at elevated breathing rates, e.g., while exercising (61 FR 25573, May 22, 1996). The available ambient monitoring data from 1988-1995 were analyzed to estimate the frequency of 5-minute peak concentrations above 500, 600, and 700 ppb, the number of repeated exceedances of these concentrations, and the sequential occurrences of peak concentrations within a given day (U.S. EPA, 1994; SAI, 1996). The analysis indicated that during that period a substantial number of 5-minute concentrations at or above 600 ppb occurred in several locations in the vicinity of certain sources (61 FR 25574, May 22, 1996). The probability of at-risk individuals breathing at elevated levels with the probability of encountering such peak concentrations was assessed in several exposure analyses (U.S. EPA, 1986, 1994; Burton et al., 1987; Rosenbaum et al., 1992; Stoeckenius et al., 1990; Sciences International, Inc., 1995).

A series of exposure analyses informed the 1994 proposed decision. These analyses focused on exposures of interest associated with coal-fired power utilities, all power utility boilers, non-utility sources of SO_2 emissions and such exposures associated with the projected reduction in emissions from fossil-fueled power plants following implementation of the acid deposition provisions (Title IV) of the 1990 Clean Air Act Amendments (U.S. EPA, 1986; Burton et al., 1987; Stoeckenius et al., 1990; Rosenbaum et al., 1992). Subsequent to the 1994 proposal, an additional exposure analysis of non-utility sources was submitted to the rulemaking docket (Sciences International, Inc., 1995). Together these analyses provided a range of estimates of the number of individuals with asthma and the percent of the population with asthma to be exposed to 5-minute concentrations of 500 and 600 ppb while at elevated exertion, as well as estimates of such individuals to be exposed on multiple occasions in a year. These analyses generally employed the time-activity exposure modeling approaches and underlying data that were available at that time.

Quantitative analyses performed for the review completed in 2010, and documented in the 2009 REA, included analyses of the limited then-available ambient air monitoring data for 5-minute concentrations in 40 U.S. counties and a population exposure assessment (75 FR 35520, June 22, 2010; 2009 REA). The air quality analyses provided estimates of the annual number of days that daily 5-minute maximum SO_2 concentrations at a monitor exceeded 5-minute concentrations of interest or benchmark concentrations[6] (2009 REA, Chapter 7). In the exposure-based approach, population-based estimates of human exposure were developed using an exposure model in order to account for the time people spend in different microenvironments, as well as for time spent at elevated breathing rates while exposed to peak 5-minute SO_2 concentrations (2009 REA, Chapter 8). The analyses were performed for recent ambient air concentrations (unadjusted, "as is" air quality), and with ambient air concentrations adjusted to just meet the then-existing annual and daily standards and several potential alternative standards.

The 2009 REA simulated population exposure using version 4.3 of the Air Pollutant Exposure (APEX) model, a probabilistic model that simulates the movement of individuals through time and space and estimates their exposure to a given pollutant in indoor, outdoor, and in-vehicle microenvironments.[7] The model was used to simulate population exposures in two study areas: Greene County, MO and a three-county portion of the St. Louis Metropolitan Statistical Area (MSA). The simulated population included all people with asthma, with results also presented for the subset of those who were children. Health risk was characterized by estimating, for each air quality scenario: (1) the number and percent of people with asthma exposed, while breathing at elevated rates, to 5-minute daily maximum SO_2 concentrations that exceeded the benchmark concentrations; and (2) the number and percent of exposed people with asthma estimated to experience moderate or greater lung function responses (in terms of FEV_1 and sRaw) at least once per year and the total number of such lung function responses estimated to occur per year (2009 REA, Chapter 8 and 9). An extensive analysis of variability and

[6] The benchmark concentrations are concentrations chosen to represent "exposures of potential concern" which were used in the analyses to estimate exposures and risks associated with 5-minute concentrations of SO_2 (75 FR 35527, June 22, 2010). Based on the evidence in the 2008 ISA and recommendations from the CASAC, staff concluded that it was appropriate to examine 5-minute benchmark concentrations in the range of 100-400 ppb (2009 REA, chapter 7). The comparisons of SO_2 concentrations to benchmark concentrations provided perspective on the extent to which, under various air quality scenarios, there was the potential for at-risk populations to experience SO_2 exposures that could be of concern.

[7] The APEX model is designed to account for sources of variability that affect people's exposures. It stochastically generates simulated individuals using census-derived probability distributions for demographic characteristics based on the information from the Census at the tract, block-group, or block-level (2009 REA).

characterization of uncertainty accompanied the exposure estimates (2009 REA, sections 8.11 and 9.4).

1.3 CURRENT REVIEW, CASAC ADVICE AND PUBLIC COMMENT

In preparing the planning document for this REA, we considered the scientific evidence presented in the second draft ISA (U.S. EPA, 2016) and the key science policy issues raised in the IRP (U.S. EPA, 2014). In February 2017, the REA Planning Document was released to the CASAC and made available for public comment (82 FR 11356, February 22, 2017). The EPA held a consultation with the CASAC and solicited comments on the REA Planning Document during a March 2017 public meeting at which the CASAC also reviewed the second draft ISA (82 FR 11356, February 22, 2017). The consultative advice from the CASAC and public comments were considered in developing the draft REA (U.S. EPA, 2017d), which implemented an exposure-based approach to assess population exposure and risk in three urban study areas (Fall River, MA, Indianapolis, IN, and Tulsa, OK). The draft REA was reviewed by the CASAC, along with the draft PA (82 FR 37213, August 9, 2017; 83 FR 14638, April 5, 2018). The EPA also solicited comment from the public on both documents (82 FR 43756 September 19, 2017; 82 FR 48507, October 18, 2017). Comments and advice from the CASAC, and public comment have been considered in development of this REA and the PA.

1.3.1 REA Aspects Updated Since 2009

As was also the case in the last review of the primary sulfur dioxide (SO_2) standards completed in 2010, the health effects evidence available in this review indicates that short-term exposures to SO_2 are causally linked to respiratory effects and that people with asthma are the at-risk population. Specifically, controlled human exposure studies demonstrate an increased risk of lung function decrements for people with asthma exposed while at increased breathing rates. The quantitative risk and exposure assessment presented in this REA is based on these findings. The approach to estimating health risk in this REA is similar to that in the REA conducted as part of the last review (2009 REA), which included quantitative analyses of both exposure and risk. Specifically, the 2009 REA included: analyses focused on short-term (5-minute) SO_2 concentrations; an exposure assessment designed to estimate exposures likely to be experienced by at-risk populations while at elevated breathing rates; and risk characterization utilizing two types of metrics: (1) comparisons of exposures to concentrations of potential concern (benchmark levels), and (2) lung function risk estimates.

The quantitative analyses performed for the current review and presented in this document reflect the use of several new pieces of information that address important areas of uncertainty identified in the last review. Perhaps most importantly, the REA uses an updated SO_2

ambient air monitoring dataset. Specifically, the data for 5-minute concentrations are greatly expanded with regard to both the number of monitoring locations for which hourly maximum 5-minute concentrations are available and the number for which all 5-minute values for each hour are available. Limitations in the 5-minute dataset available at the time of the last review influenced the approaches that could be used in the 2009 REA to characterize the potential for at-risk populations to experience exposures of potential concern. The analysis approach for this REA is based on linking the health effects information to population exposure estimates which draws on the improved understanding of 5-minute SO_2 concentrations and takes advantage of a number of improvements and updates to the air quality, exposure, and risk models, and their associated input data.

As in the last review, this REA uses the Air Pollutant Exposure model (APEX) to estimate population exposures that account for the time people spend in different microenvironments, as well as for time spent at elevated breathing rates while exposed to peak 5-minute SO_2 concentrations. The REA also reflects the new information and model improvements that are now available including:

- A SO_2 air monitoring dataset that is greatly expanded with regard to both the number of monitoring locations for which hourly maximum 5-minute concentrations are available and the number for which all 5-minute values for each hour are available;
- Estimated exposures associated with air quality adjusted to just meet the current standard across a three-year averaging period;[8]
- Improvements in the air quality dispersion model, AERMOD, intended to reduce uncertainties in 1-hour concentration estimates;
- Greatly expanded database of human activity patterns that provide a stronger foundation for inhalation exposure modeling;
- Improvements to the exposure model, APEX, designed to reduce uncertainties in personal attributes of simulated individuals (e.g., breathing rates); and,
- Use of an expanded dataset for development of a lung function exposure-response function, intended to reduce uncertainties in the response across the range of the study data.

Based on the new information, model improvements, exploratory data evaluations, and updated characterization of uncertainties, the results from this REA provide an improved characterization of exposure and risk to inform the EPA's review of the primary SO_2 standard.

[8] The 2009 REA estimated exposures considering air quality adjusted to just meet several alternative standards across a single-year period.

The exposure-based risk assessment for this review includes an assessment of air quality conditions just meeting the current standard in three study areas. The study areas were selected based on consideration of the magnitude of recent SO_2 concentrations, number of monitors in the area, including those with 5-minute monitoring data, and population size. The risk characterization is based on comparisons of population 5-minute exposures at elevated breathing rates to health-based benchmark levels and estimated population risk of moderate or greater SO_2-related lung function decrements. The analyses and results are documented in this REA and key findings of the REA are considered in the broader context of the PA, which also considers the current evidence as assessed in the ISA and characterization of SO_2 concentrations in ambient air across the U.S. based on recent monitoring data, with particular attention to peak 5-minute concentrations.

1.3.2 CASAC Advice and Public Comment

After consultation with the CASAC on the REA Planning Document (U.S. EPA, 2017c) in March 2017 (Diez Roux, 2017), the EPA developed the draft REA (U.S. EPA, 2017d). The CASAC SO_X Panel discussed its review of the draft REA at a public meeting on September 20-21, 2017 and in a public teleconference on April 20, 2018. The CASAC comments and recommendations on the draft REA are provided in a May 2018 letter to the Administrator (Cox and Diez Roux, 2018a). A number of comments on aspects of the draft REA were also received from the public (see the public docket for this review, EPA-HQ-OAR-2013-0566 at www.regulations.gov).

This final REA has been produced in consideration of the comments received on the draft REA from the CASAC and from the public. The approach used to estimate population exposure and risk has remained largely the same as the approach used in the draft REA, with a number of adjustments and additions to address comments. Key changes include:

- Clarification regarding key design aspects including the air quality scenario and scope of REA (Chapter 2);
- Revised study area maps that show locations of meteorological stations, air quality receptors, emissions sources, and ambient air monitors, that also indicate source types and SO_2 emissions (sections 3.2 and 3.4);
- Improvements in estimating ambient concentrations associated with sources not explicitly modeled in the Indianapolis study area (section 3.2.4);
- Additional evaluations of the daytime estimated 1-hour and 5-minute ambient air concentrations in the three study areas by season (sections 3.2.5 and section 3.5.3.3);
- Expanded discussion regarding the approach used to adjust air quality to just meet the current standard (section 3.4);

- Use of newly acquired continuous 5-minute ambient air monitoring data to estimate 5-minute concentrations at modeled air quality receptors in the Indianapolis study area (section 3.5.1);
- Analysis of asthma prevalence information regarding the influence of body mass index and race on populations with asthma (section 4.1.2);
- Reorganized, clarified and expanded discussion regarding exposure model input data (i.e., body weight, surface area, energy expenditure) and algorithms (i.e., resting metabolic rate, breathing rate) (section 4.1.3);
- Expanded discussion of using activity pattern data from any individual in CHAD, regardless of whether their asthma status is known or unknown, to represent the simulated individuals with asthma (section 4.3.3);
- Additional analysis and revised study area maps to better indicate where study area populations overlap with highest ambient SO_2 concentrations (sections 5.1 and 5.4); and
- Updated analysis of the microenvironments where the simulated population experiences the highest exposures (section 5.2);
- Inclusion of number (and percentage) of individuals in the estimates of population exposure and risk of lung function decrements presented in summary tables (section 5.2 and 5.3);
- Expanded discussion of previously identified uncertainties, as well as identification and discussion of additional uncertainties (Table 6-3).

REFERENCES

Burton CS, Stoeckenius TE, Stocking TS, Carr EL, Austin BS, Roberson RL (1987). Assessment of Exposures of Exercising Asthmatics to Short-term SO_2 Levels as a Result of Emissions from U.S. Fossil-fueled Power Plants. Systems Applications Inc., San Rafael, CA. Publication No. 87/176, September 23, 1987.

Cox, LA; Diez Roux, A. (2018a). Letter from Louis Anthony Cox, Chair, Clean Air Scientific Advisory Committee, and Ana Diez Roux, Immediate Past Chair, Clean Air Scientific Advisory Committee, to Administrator E. Scott Pruitt. Re: CASAC Review of the EPA's Risk and Exposure Assessment for the Review of the Primary National Ambient Air Quality Standard for Sulfur Oxides (External Review Draft - August 2017). April 30, 2018.

Cox, LA; Diez Roux, A. (2018b). Letter from Louis Anthony Cox, Chair, Clean Air Scientific Advisory Committee, and Ana Diez Roux, Immediate Past Chair, Clean Air Scientific Advisory Committee, to Administrator E. Scott Pruitt. Re: CASAC Review of the EPA's Policy Assessment for the Review of the Primary National Ambient Air Quality Standard for Sulfur Oxides (External Review Draft - August 2017). April 30, 2018.

Diez Roux A. (2017). Letter from Ana Diez Roux, Chair, Clean Air Scientific Advisory Committee, to Administrator Gina McCarthy. Re: Consultation on the EPA's Review of the Primary National Ambient Air Quality Standard for Sulfur Oxides: Risk and Exposure Assessment Planning Document (External Review Draft – February 2017). April 18, 2017.

Rosenbaum AS, Hudischewskyj AB, Roberson RL, Burton CS. (1992). Estimates of Future Exposures of Exercising Asthmatics to Short-term Elevated SO_2 Concentrations Resulting from Emissions of U.S. Fossil-fueled Power Plants: Effects of the 1990 Amendments to the Clean Air Act and a 5-Minute Average Ambient SO2 Standard. Publication No. SYSAPP- 92/016. April 23, 1992. Docket No. A-84-25, IV-K-37.

SAI. (1996). Summary of 1988-1995 Ambient 5-Minute SO_2 Concentration Data. Prepared for US EPA Office of Air Quality Planning and Standards by Systems Application International. Contract #68-D3-0101, May 1996.

Sciences International (1995). Estimate of the Nationwide Exercising Asthmatic Exposure Frequency to Short-term Peak Sulfur Dioxide Concentrations in the Vicinity of Non-Utility Sources. Prepared for National Mining Association by Sciences International, Inc., Alexandria VA. April 1995. Docket No. A-84-25, VIII-D-71.

Stoeckenius TE, Garelick B, Austin BS, O'Connor K, Pehling JR. (1990). Estimates of Nationwide Asthmatic Exposures to Short-Term Sulfur Dioxide Concentrations in the Vicinity of Non-Utility Sources. Systems Applications Inc., San Rafael, CA. Publication No. SYSAPP-90/129, December 6, 1990.

U.S. EPA. (1986). Review of the National Ambient Air Quality Standards for Sulfur Oxides: Updated Assessment of Scientific and Technical Information, Addendum to the 1982 OAQPS Staff Paper. Office of Air Quality Planning and Standards, Research Triangle Park, NC, EPA/450/5-86-13. Available from: NTIS, Springfield, VA; PB87-200259/XAB.

U.S. EPA. (1994). Supplement to the second addendum (1986) to air quality criteria for particulate matter and sulfur oxides (1982): Assessment of new findings on sulfur dioxide acute exposure health effects in asthmatic individuals. Environmental Criteria and Assessment Office, Office of Health and Environmental Assessment, Office of Research and Development, Research Triangle Park, NC, EPA/600/FP-93/002, August 1994. Available at: https://www3.epa.gov/ttn/naaqs/standards/so2/s_so2_pr.html

U.S. EPA. (2009). Risk and Exposure Assessment to Support the Review of the SO_2 Primary National Ambient Air Quality Standard. Office of Air Quality Planning and Standards, Research Triangle Park, NC, EPA-452/R-09-007, July 2009. Available at: https://www3.epa.gov/ttn/naaqs/standards/so2/data/200908SO2REAFinalReport.pdf

U.S. EPA. (2014). Integrated Review Plan for the Primary National Ambient Air Quality Standard for Sulfur Dioxide, Final. Office of Air Quality Planning and Standards, Research Triangle Park, NC, EPA-452/P-14-007, October 2014. Available at: https://www3.epa.gov/ttn/naaqs/standards/so2/data/20141028so2reviewplan.pdf

U.S. EPA. (2016). Integrated Science Assessment (ISA) for Sulfur Oxides – Health Criteria (Second External Review Draft). National Center for Environmental Assessment-RTP Division, Office of Research and Development, Research Triangle Park, NC, EPA/600/R-16/351, December 2016. Available at: https://cfpub.epa.gov/ncea/isa/recordisplay.cfm?deid=326450

U.S. EPA. (2017a). Integrated Science Assessment (ISA) for Sulfur Oxides – Health Criteria (Final). National Center for Environmental Assessment-RTP Division, Office of Research and Development, Research Triangle Park, NC, EPA/600/R-17/451, December 2017. Available at: https://cfpub.epa.gov/ncea/isa/recordisplay.cfm?deid=338596

U.S. EPA. (2017b). Integrated Review Plan for the Secondary National Ambient Air Quality Standard for Ecological Effects of Oxides of Nitrogen, Oxides of Sulfur and Particulate Matter. Office of Air Quality Planning and Standards, Research Triangle Park, NC, EPA-452/R-17-002, January 2017. Available at: https://www.epa.gov/naaqs/nitrogen-dioxide-no2-and-sulfur-dioxide-so2-secondarystandards-planning-documents-current

U.S. EPA. (2017c). Review of the Primary National Ambient Air Quality Standard for Sulfur Oxides: Risk and Exposure Assessment Planning Document. EPA-452/P-17-001, February 2017. Available at: https://www3.epa.gov/ttn/naaqs/standards/so2/data/20170216so2rea.pdf

U.S. EPA. (2017d). Review of the Primary National Ambient Air Quality Standard for Sulfur Oxides: Risk and Exposure Assessment Planning Document. Office of Air Quality Planning and Standards, Research Triangle Park, NC, EPA-452/P-17-001, February

2017. Available at:
https://www3.epa.gov/ttn/naaqs/standards/so2/data/20170216so2rea.pdf

U.S. EPA. (2017e). Policy Assessment for the Review of the Primary National Ambient Air Quality Standard for Sulfur Oxides, External Review Draft. Office of Air Quality Planning and Standards, Research Triangle Park, NC, EPA-452/P-17-003, August 2017. Available at: https://www.epa.gov/naaqs/sulfur-dioxide-so2-primary-air-quality-standards

U.S. EPA. (2018). Policy Assessment for the Review of the Primary National Ambient Air Quality Standard for Sulfur Oxides, Final. Office of Air Quality Planning and Standards, Research Triangle Park, NC, EPA-452/R-18-002, May 2018. Available at: https://www.epa.gov/naaqs/sulfur-dioxide-so2-primary-air-quality-standards

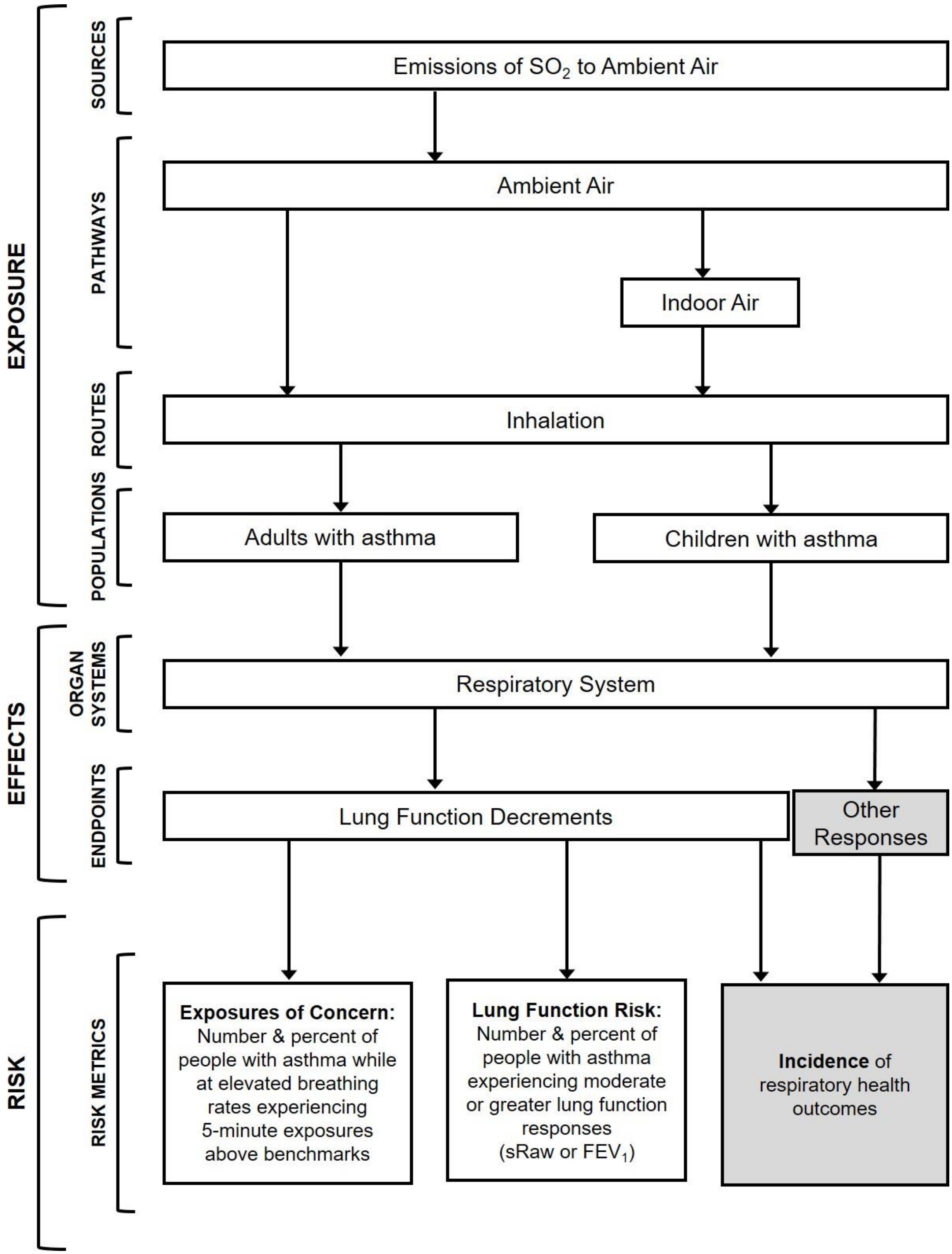
2 OVERVIEW OF ASSESSMENT APPROACH

As summarized in the IRP and PA for this review of the NAAQS for SO_X, the review focuses on the presence in ambient air of sulfur oxides, a group of closely related gaseous compounds that include sulfur dioxide and sulfur trioxide and of which sulfur dioxide (the indicator for the current standard) is the most prevalent in the atmosphere and the one for which there is a large body of scientific evidence on health effects. Sulfur trioxide is known to be present in the emissions of coal-fired power plants, factories, and refineries, but it reacts with water vapor within emission stacks or immediately after release into the atmosphere within seconds to form H_2SO_4 which quickly condenses onto existing atmospheric particles or participates in new particle formation (ISA, p. 2-18). Thus, only SO_2 is present at concentrations in the gas phase that are relevant for chemistry in the atmospheric boundary layer and troposphere, and for human exposures (ISA, p. 2-18). The health effects of particulate atmospheric transformation products of SO_X, such as sulfates, are addressed in the review of the NAAQS for particulate matter (U.S EPA, 2018; U.S. EPA, 2016). For these reasons, this REA is focused on SO_2.[9] The conceptual model for exposure and associated health risk of SO_2 in ambient air that guides our assessment in this review is described in this section along with an overview of the implemented approach.

2.1 CONCEPTUAL MODEL FOR SO_2 EXPOSURE AND RISK

The conceptual model for our consideration of exposure and risk associated with SO_2 in ambient air is illustrated in Figure 2-1. This general model guided our assessment in the last review and, as discussed in the REA Planning Document and draft REA, remains appropriate in the current review. The unshaded boxes indicate components included in the assessment in this review. Current information regarding the individual components specified in the model (emissions sources, exposure pathways, routes of exposure, exposed populations, health endpoints and risk metrics) is summarized in the following sections. A more detailed characterization of this information is presented in the ISA (U.S. EPA, 2017a).

[9] While there are some toxicological animal studies of SO_X in mixtures, such as with co-occurring PM, that indicate some enhanced effect on lung function parameters, there are a number of limitations with regard to appropriate controls and relevance to ambient air exposures (ISA, pp. 5-143 to 5-144). Thus, the available information does not support characterization in this assessment of any potential for modification SO_2-related effects by copollutants, such as PM. Uncertainties in the exposure and risk estimates generated in this REA with regard to the potential for modification of SO_2-related effects by co-occurring pollutants, such as PM, are characterized in section 6.2.1.



Note: The grey boxes indicate elements not included.

Figure 2-1. Conceptual model for exposure and associated health risk of SO_2 in ambient air.

2.1.1 Sources of SO_2

Sulfur dioxide occurs in ambient air as a result of direct emissions of SO_2 as well as emissions of other compounds, such as reduced sulfur compounds or sulfides, that are converted to SO_2 through chemical reactions in the atmosphere. The largest natural sources of SO_2 are volcanoes and wildfires. Fossil fuel combustion is the main anthropogenic source of SO_2 and industrial chemical production and pulp and paper production are among the sources of reduced sulfur compounds that are converted to SO_2 in the atmosphere. Anthropogenic sources of SO_2 emissions that contribute to SO_2 found in ambient air are primarily large facilities and include coal-fired electricity generating units (EGUs) as well as other industrial facilities (U.S. EPA, 2008 [hereafter referred to as the 2008 ISA], section 2.1; ISA, section 2.2.1). Because such large, discrete sources are the primary source of SO_2 (e.g., versus more prevalent, widespread sources), ambient concentrations can vary substantially across an area and can be relatively high in areas affected by these large sources.

Coal-fired EGUs are an important emissions source because coal contains sulfur, which is present to some degree in all fossil fuels. The sulfur content of the most common types of coal varies between 0.4 and 4% by mass (ISA, section 2.2). Fuel sulfur is almost entirely converted to sulfur oxides during combustion. This makes accurate estimates of SO_2 combustion emissions possible based on fuel composition and combustion rates (ISA, section 2.2). Fuel combustion by electric utilities as well as industrial and other sources is the largest source of anthropogenic SO_2 emissions (ISA, Figure 2-1).

Although they may be fewer in number than fossil fuel-fired EGUs, other types of large emissions facilities that may impact local air quality include copper smelters, kraft pulp mills, Portland Cement plants, iron and steel mill plants, sulfuric acid plants, petroleum refineries, and chemical processing plants. For example, the metal processing sector represents less than 2.3% of total emissions from the 2014 National Emissions Inventory (NEI),[10] however, monitoring sites that have recorded some of the highest 1-hour daily maximum SO_2 concentrations in the U.S. are located near copper smelters in Arizona (ISA, sections 2.5.2 and 2.5.4, Figure 2-11). The two smelters in this area emit appreciable quantities of SO_2, estimated at 17,000 tons per year (tpy) and 5,000 tpy (ISA, p. 2-50), but for added perspective, several EGUs in other areas have been estimated to emit well over 50,000 tpy in the 2014 NEI.

The main indoor source of SO_2 is indoor combustion of sulfur-containing fuels, such as from space heaters that are generally used in the U.S. as emergency or supplemental sources of heat. For example, a study in the eastern U.S. reported that kerosene heaters, but not fireplaces,

[10] The National Emissions Inventory (NEI) is a comprehensive and detailed estimate of air emissions of criteria pollutants, criteria precursors, and hazardous air pollutants from air emissions sources. For additional information, see https://www.epa.gov/air-emissions-inventories/national-emissions-inventory-nei.

woodstoves, or gas space heaters, resulted in increased indoor concentrations of SO_2 (ISA, section 3.4.1.1). Personal SO_2 exposure measurements, however, have generally been lower than ambient air concentrations, indicating personal exposure is generally dominated by ambient air (outdoor) sources (ISA, section 3.4.1).

The context for the REA is exposure and associated risk of SO_2 emitted into ambient air. Accordingly, the conceptual model for the REA focuses on sources to ambient air (Figure 2-1).

2.1.2 Exposure Pathways and Route

Human exposure to SO_2 involves the contact between a person and the pollutant in any of the various locations (or microenvironments, MEs) in which people spend their time. As SO_2 is a gas, human exposure occurs through inhalation of air containing SO_2. The concentrations of SO_2 occurring in each ME and the associated activity performed in the ME at the time of exposure both contribute to individual exposure events. Together, these exposure events make up an individual's exposure (ISA, section 3.2.2).

Exposure microenvironments occur indoors (e.g., in homes, offices or stores), outdoors (e.g., yards, parks, sidewalks) and in vehicles (e.g., automobiles, buses). All of these microenvironments can receive ambient air that may contain SO_2. Thus, the pathways by which people are exposed to SO_2 in ambient air involve inhaling air while spending time in the various MEs.

While indoors, people can be exposed to SO_2 from indoor sources as well as to SO_2 associated with outdoor air that has infiltrated into indoor MEs. Studies of personal exposure have generally found that the largest portion of a person's day is spent indoors (ISA, section 3.4.2.1). As a result of this and indoor SO_2 concentrations typically being lower than SO_2 concentrations measured outdoors, SO_2 exposure concentrations are often much lower than SO_2 concentrations in ambient air (ISA, section 3.4.1). As stated in the ISA, high correlations (>0.75) between indoor and outdoor SO_2 concentrations indicate that variations in outdoor ambient SO_2 concentration[11] are driving indoor SO_2 concentrations, which is considered to be consistent with the relative lack of indoor sources of SO_2 (ISA, section 3.4.1.2).

Thus, personal SO_2 exposure is expected to be dominated by SO_2 emitted into ambient air in outdoor microenvironments and enclosed microenvironments with high air exchange rates, such as buildings with open windows and vehicles. This was found to be the case in exposure

[11] Concentrations of SO_2 in ambient air are spatially highly variable compared to pollutants such as ozone (ISA, section 3.2.3); this is due to the point source nature of SO_2 emissions. Another factor in the spatial variability is the dispersion and oxidation of SO_2 in the atmosphere, processes that contribute to decreasing concentrations with increasing distance from the source. Point source emissions of SO_X create a plume of higher concentrations, which may or may not impact large portions of surrounding populated areas depending on meteorological conditions and terrain (ISA, section 3.2.3).

modeling of recent air quality performed for the 2009 REA: more than 80% of the events by which simulated individuals experienced elevated 5-minute exposure concentrations of interest were in outdoor MEs (2009 REA, Figure 8-21). As was done in the 2009 REA for the last review of the NAAQS for SO_X, exposures to SO_2 in ambient air outdoors, as well as to ambient air that has infiltrated indoors, are included in the REA for the current review.

2.1.3 At-Risk Populations

As at the time of the 2009 REA, the current evidence demonstrates that the populations at increased risk of effects from SO_2 exposure continue to be people with asthma, particularly children with asthma (ISA, section 6.3.1). Strong evidence of this comes from the controlled human exposure studies of people with asthma exposed to SO_2 when their breathing rates are increased, such as from exercise (ISA, section 5.2.1.9). Consistent with the controlled human exposure study findings of asthma exacerbation-related effects, some epidemiological studies in the current evidence report associations between short-term SO_2 exposure and increased risk of asthma-related emergency department visits and hospital admissions (ISA, section 5.2.1.9).

The short-term respiratory effects that are the focus of the quantitative assessment, and for which the evidence for respiratory effects associated with policy-relevant SO_2 exposure concentrations is strongest, are asthma exacerbation-related effects (ISA, Table 1-1). Under resting conditions, inhaled SO_2 is readily removed in the nasal passages (ISA, section 1.5.1). However, during activities that result in elevated breathing rates, such as those associated with exercise, and/or an increased potential for taking breaths through the mouth (versus the nose), there is greater transport of inhaled SO_2 past the nasal passages to the tracheobronchial region of the airways where it can contribute to bronchoconstriction-related effects and asthma exacerbation (ISA, section 1.5.1). Thus, elevated breathing rates and breathing habits that include breathing through the mouth (oronasal), such as that occurring during exercise, play important roles in eliciting SO_2-related effects in at-risk populations.

While some controlled exposure studies involving adolescents with asthma have indicated that this age group has similar responsiveness as adults, controlled exposure study data are not available for children younger than 12 years (ISA, section 5.2.1.2). However, some factors indicate that children (e.g., younger than 13 years) with asthma may be at a greater risk than adults with asthma. For example, children, particularly younger than 13 years of age, have a greater tendency to breathe through the mouth than do adults (ISA, section 4.1.2.2). Evidence also suggests that older adults with asthma may also be at an increased risk compared to younger adults with asthma (ISA, section 6.5.1.2).

The evidence in controlled exposure studies documents the difference in sensitivity to SO_2-related respiratory effects in individuals with and without asthma. For example, these studies document respiratory effects occurring in exercising study subjects with asthma at exposure concentrations below 1000 ppb, while higher concentrations are needed to elicit similar effects in healthy subjects and in some subjects with asthma (ISA, sections 5.2.1.2 and 5.2.1.7).[12] The currently available information does not identify other populations at increased risk beyond what is described here (ISA, section 6.6). As indicated in Figure 2-1, people with asthma, including both adults and children, are specifically identified as at-risk populations in the REA for this review.

2.1.4 Health Endpoints

The health effects causally related to SO_2 exposures are effects on the respiratory system (ISA, section 1.6). As demonstrated in long-standing evidence from controlled human exposure studies and consistent with findings in epidemiological studies, short-term SO_2 exposures (as short as a few minutes) can result in asthma exacerbation-related effects in people with asthma. The controlled human exposure studies have demonstrated a relationship between 5- and 10-minute peak SO_2 exposures and bronchoconstriction-related decrements in lung function in exercising individuals with asthma; depending on the exposure level, these decrements are accompanied by respiratory symptoms (ISA, section 5.2.1.2).

Lung function decrements were quantified in these studies by reductions in forced expiratory volume in one second, FEV_1, and increased specific airway resistance, sRaw. In considering the magnitude of these responses, the ISA (as in the 2008 ISA) focuses on 15% or greater reductions in FEV_1 and increases in sRaw of 100% or more (ISA, sections 1.6.1.1 and 5.2.1.2). Such responses have been reported in some individuals with asthma exposed to 5-minute concentrations as low as 200 ppb while exercising. Across the range of exposure concentrations studied, both the percentage of individuals affected to at least this degree and the severity of the response increases with increasing SO_2 concentrations. At higher concentrations (above 400 ppb), such responses were frequently accompanied by respiratory symptoms (ISA, section 5.2.1.2).

2.1.5 Risk Metrics

As was the case in the 2009 REA, the risk metrics included in the current REA (bottom panels, Figure 2-1) are based on the SO_2-induced bronchoconstriction-related lung function

[12] The evidence from controlled exposure studies has long documented the sizeable variation in sensitivity to SO_2 among individuals with asthma. This was further characterized in a pooled analysis of data from five such studies that is newly available in this review (Johns et al., 2010). This new analysis demonstrates the study population of individuals with asthma to fall into one of two subpopulations with regard to airway responsiveness to SO_2. One subpopulation is insensitive to the bronchoconstrictive effects of SO_2 even at concentrations as high as 1.0 ppm, and it is the second subpopulation that has an increased risk for bronchoconstriction at the lower concentrations of SO_2 (ISA, section 5.2.1.2).

decrements documented in the strong evidence base of controlled human exposure studies of exercising individuals with asthma. Bronchoconstriction, an asthma-exacerbation-related effect, is the “most sensitive indicator of SO_2-induced lung function effects” and the evidence for this effect is strong (ISA, section 5.2.1.2, p. 5-8). The first of the risk metrics included in this REA involves characterization of the extent to which individuals with asthma were estimated to experience 5-minute exposures at or above concentrations of potential concern while they are at elevated breathing rates. The second metric quantifies the extent to which individuals with asthma are estimated to experience lung function responses (in terms of a doubling, or larger increase, in sRaw) as a result of 5-minute SO_2 exposures while at elevated breathing rates.

In deriving these two risk metrics, the controlled human exposure studies are used in two ways: (1) to identify exposure concentrations of potential concern (“benchmark concentrations”) and (2) to derive exposure-response (E-R) functions for lung function decrements. As described in more detail in section 3.5.1, the benchmark concentrations are 5-minute exposure concentrations chosen to represent exposures of potential concern. The first metric, the comparison of SO_2 exposures to benchmark concentrations, provides perspective on the extent to which there is potential for sensitive individuals with asthma to experience SO_2 exposures that could be of concern at air quality just meeting the current standard.

The second metric relies on the E-R function and exposure estimates to estimate risk of decrements in lung function based on sRaw, which is a specific measure of bronchoconstriction. The focus on sRaw as the primary indicator of lung function response is consistent with the emphasis on this indicator in the REA for the last review. The E-R functions for sRaw are based on more observations from individual subjects than were E-R functions based on FEV_1 (2009 REA, p. 332), which provides greater confidence in the resultant quantitative relationship when compared with that developed for the FEV_1 health endpoint.

Another category of metric shown in the conceptual model figure represents potential asthma-exacerbation-related health outcomes that are reported in the epidemiological evidence. As indicated by the shading in Figure 2-1, this category of metrics is not included in this REA as the current evidence base does not support its inclusion. This was also the case in the 2009 REA (REA Planning Document, section 3.2.3). As examined in detail in the ISA, the epidemiological evidence includes studies reporting associations between short-term SO_2 concentrations and asthma-related emergency department visits or hospitalizations. The risk characterization for the 2009 REA focused on metrics for lung function decrements related to bronchoconstriction, concluding that the epidemiological evidence did not support development of an epidemiological study-based risk model. In considering support in the evidence available in this review, the REA Planning Document for this REA reached the same conclusion (REA Planning Document, section 3.2.3). Thus, as shown in Figure 2-1, the incidence of respiratory health outcomes metric is not included in this REA.

2.2 ASSESSMENT APPROACH

The approach employed for this REA generally involves estimating population exposures to ambient air-related SO_2 concentrations and associated health risk for air quality conditions simulated to just meet the current standard (Figure 2-2). This approach, which draws on air monitoring data, air quality modeling and exposure modeling, was applied in three study areas (section 3.1) selected to be most informative to this review. The focus on air quality conditions just meeting the current standard reflects the key overarching question articulated in the IRP for this review: Does the currently available scientific evidence- and exposure/risk-based information, as reflected in the ISA and REA, support or call into question the adequacy of the protection afforded by the current standard (IRP, section 3; PA, section 3.2)? In considering the final ISA and the draft REA results, the draft PA reached preliminary conclusions that the answer to this question was no and that it is appropriate to consider retaining the current standard without revision. The CASAC concurred with this conclusion (Cox and Diez Roux, 2018). Accordingly, exposure and risk analyses using alternative air quality conditions were not warranted and have not been performed for this REA.

As indicated by the case study approach, the REA analyses are not intended to provide a comprehensive national assessment. Rather, they are intended to provide assessments for a small varied set of study areas, and the associated exposed at-risk populations, that will be informative to EPA's consideration of potential exposures and risks that may be associated with the air quality conditions occurring under the current SO_2 standard. The purpose of the REA is to assess, based on the currently available, improved and expanded tools and information, the potential for exposures and risks beyond those indicated by the information available at the time the current standard was established. In this way, the REA can inform the EPA's conclusions on the public health protection afforded by the current standard.

Consistent with the health effects evidence and the health risk metrics identified in section 2.1.5, the focus is on short-term exposures of individuals in the population with asthma during times when they are breathing at an elevated rate. Exposure and risk is characterized for two population groups: adults (individuals older than 18 years) with asthma and school-aged children (aged 5 to 18 years)[13] with asthma. The focus on these populations is consistent with the ISA's identification of individuals with asthma as the population at risk of SO_2-related effects,

[13] As in other NAAQS reviews, this REA does not estimate exposures and risk for children younger than 5 years old due to the more limited information contributing relatively greater uncertainty in modeling their activity patterns and physiological processes than children between the ages of 5 to 18.

and its conclusion that within this population, children with asthma may be at greater risk than adults with asthma (ISA, section 6.6).

In order to estimate ambient air concentrations at the needed temporal scale of 5-minute increments, the REA employs air quality modeling as informed by additional information from 5-minute ambient air monitoring data. Air quality modeling is used in order to adequately capture the spatial variation in ambient SO_2 concentrations across an urban area, which can be relatively high in areas affected by large point sources and which the limited number of monitoring locations in each area are unlikely to capture. Continuous 5-minute ambient air monitoring data are used to reflect the fine-scale temporal variation in SO_2 concentrations documented by these data and for which air quality modeling is limited, e.g., by limitations in currently available input data such as emissions estimates. Thus, 5-minute concentrations in ambient air were estimated using a combination of 1-hour concentrations from the EPA's preferred near-field dispersion model, the American Meteorological Society/EPA regulatory model (AERMOD), and relationships between 1-hour and 5-minute concentrations occurring in the local ambient air monitoring data.[14]

The Air Pollutants Exposure (APEX) model, a probabilistic human exposure model that simulates the activity of individuals in the population, including their exertion levels and movement through time and space, was then used to estimate 5-minute exposure concentrations for individuals based on exposures in indoor, outdoor, and in-vehicle microenvironments. The use of APEX for estimating exposures allows for consideration of factors that affect exposures that are not addressed by consideration of ambient air concentrations alone. These factors include 1) attenuation in SO_2 concentrations expected to occur in some indoor microenvironments, 2) the influence of human activity patterns on the time series of exposure concentrations, and 3) accounting for human physiology and the occurrence of elevated breathing rates concurrent with SO_2 exposures, all key to appropriately characterizing health risk for SO_2.

The estimated exposures were then combined with findings of the controlled human exposure studies to characterize health risk using two approaches. The first approach compares estimated exposures to benchmark concentrations of interest and the second combines exposures with an E-R function to estimate the expected occurrences of decrements in lung function.

---

[14] The current information continues to support the use of an air dispersion model such as AERMOD over the use of other models, such as photochemical models, for modeling of directly emitted SO_2 concentrations for use in assessing risk and exposure for this pollutant. Unlike dispersion models, photochemical models cannot capture the sharp concentration gradients that can occur near SO_2 sources. Also, SO_2 emissions to ambient air are dominated by point sources, such as large coal-fired utilities, and AERMOD is the EPA's preferred air quality model for SO_2 for State Implementation Plans (SIPs) and new source permitting purposes. For all of these reasons, AERMOD remains the most appropriate model for predicting SO_2 concentrations in ambient air.

Thus, two types of risk metrics were derived from the simulated individual exposure profiles: (1) the number and percent of the simulated subpopulation that had at least one 5-minute exposure above the benchmark concentrations of 100, 200, 300, and 400 ppb and (2) the number and percent per year of simulated at-risk individuals that would experience moderate or greater lung function decrements in response to 5-minute daily maximum peak exposures while engaged in moderate or greater exertion. Estimates were developed for three study areas. The details and basis for each of these aspects of the assessment are described in Chapters 3 and 4.

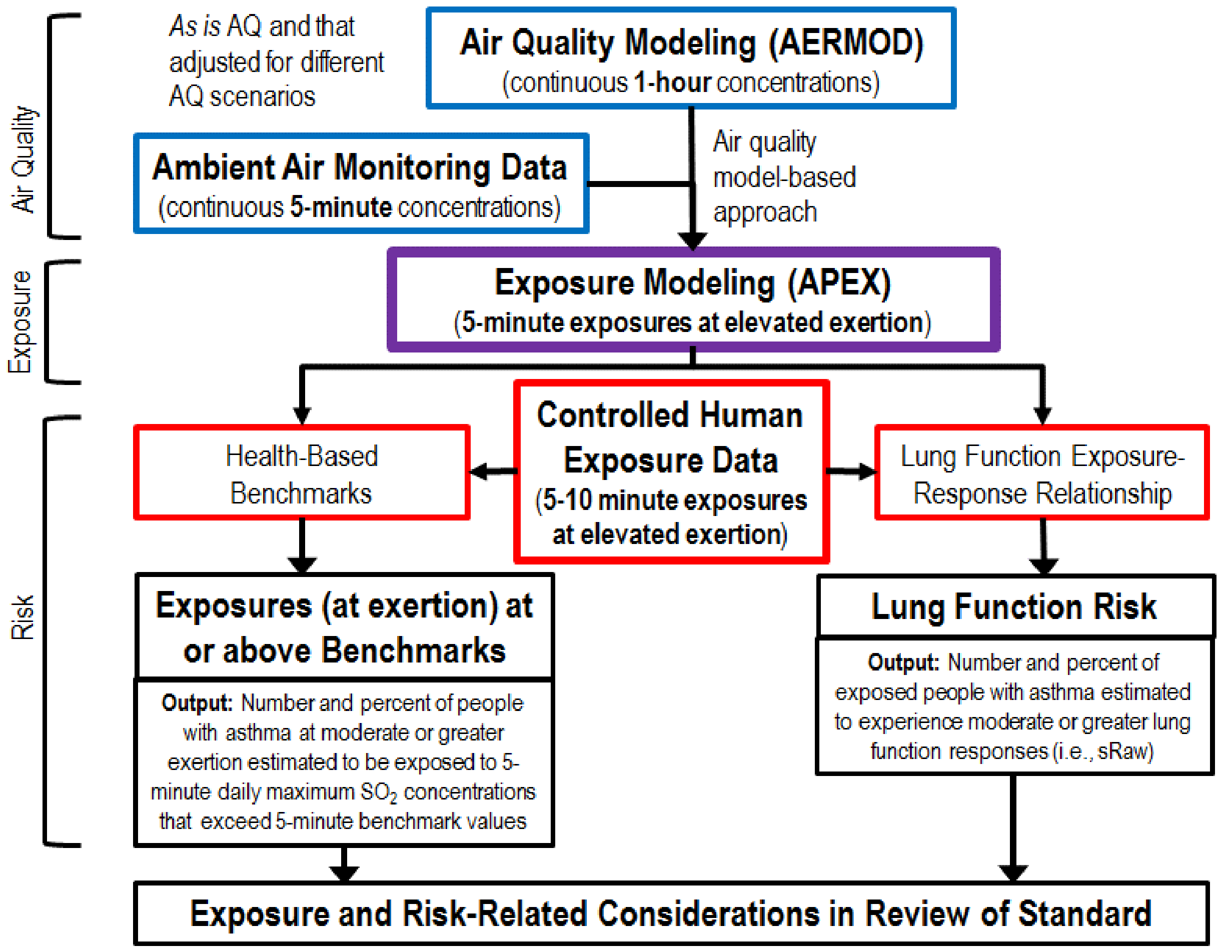


Figure 2-2. Overview of the assessment approach.

REFERENCES

Cox, LA; Diez Roux, A. (2018). Letter from Louis Anthony Cox, Chair, Clean Air Scientific Advisory Committee, and Ana Diez Roux, Immediate Past Chair, Clean Air Scientific Advisory Committee, to Administrator E. Scott Pruitt. Re: CASAC Review of the EPA's Policy Assessment for the Review of the Primary National Ambient Air Quality Standard for Sulfur Oxides (External Review Draft - August 2017). April 30, 2018.

Johns DO, Svendsgaard D, Linn WS. (2010). Analysis of the concentration-respiratory response among asthmatics following controlled short-term exposures to sulfur dioxide. Inhal Toxicol. 22:1184-1193. http://dx.doi.org/10.3109/08958378.2010.535220

U.S. EPA. (2008). Integrated Science Assessment (ISA) for Sulfur Oxides – Health Criteria (Final Report). National Center for Environmental Assessment-RTP Division, Office of Research and Development, Research Triangle Park, NC, EPA-600/R-08/047F, September 2008. Available at: http://cfpub.epa.gov/ncea/cfm/recordisplay.cfm?deid=198843

U.S. EPA. (2009). Risk and Exposure Assessment to Support the Review of the SO_2 Primary National Ambient Air Quality Standard. Office of Air Quality Planning and Standards, Research Triangle Park, NC, EPA-452/R-09-007, July 2009. Available at: https://www3.epa.gov/ttn/naaqs/standards/so2/data/200908SO2REAFinalReport.pdf

U.S. EPA. (2016). Integrated Review Plan for the Secondary National Ambient Air Quality Standards for Particulate Matter. Office of Air Quality Planning and Standards, Research Triangle Park, NC, EPA-452/R-16-005, December 2016. Available at: https://www.epa.gov/naaqs/particulate-matter-pm-standards-planning-documents-current-review

U.S. EPA. (2017). Integrated Science Assessment (ISA) for Sulfur Oxides – Health Criteria (Final). National Center for Environmental Assessment-RTP Division, Office of Research and Development, Research Triangle Park, NC, EPA/600/R-17/451, December 2017. Available at: https://cfpub.epa.gov/ncea/isa/recordisplay.cfm?deid=338596

U.S. EPA. (2018). Policy Assessment for the Review of the Primary National Ambient Air Quality Standard for Sulfur Oxides, Final. Office of Air Quality Planning and Standards, Research Triangle Park, NC, EPA-452/R-18-002, May 2018. Available at: https://www.epa.gov/naaqs/sulfur-dioxide-so2-primary-air-quality-standards

3 AMBIENT AIR CONCENTRATIONS

As summarized in chapter 2, the approach for this REA is based on linking the health effects information to estimated population-based exposures that reflect our current understanding of 5-minute concentrations of SO_2 in the ambient air. This approach is applied to three study areas to provide a valuable perspective on exposures and risks for at-risk populations that is informative to this review of the SO_2 primary standard. This chapter describes the methodology for developing the spatial and temporal patterns of 5-minute concentrations in ambient air for each of the three study areas. Our overall objective for this methodology is not necessarily to develop an air quality surface for each study area that exactly matches one that has occurred. Rather, it is to develop a hypothetical air quality scenario that accounts for the spatial and temporal pattern of ambient SO_2 concentrations in each study area that might be expected to occur when the current primary SO_2 standard has just been met, is based on the types of SO_2 sources that have existed in the area (and local or nearby sources that may also influence ambient air concentrations) and considers the expected variability in observed meteorological conditions. This hypothetical scenario, however, is not necessarily reflective of a specific calendar year, even though data from specific years have been used as a basis for the development of the hypothetical scenario. In so doing, we have implemented methods intended to capture the appropriate spatial and temporal heterogeneity in SO_2 concentrations that occur near and around important emissions sources considering this hypothetical air quality scenario and, when considering population demographics, to reasonably represent the population groups at risk for SO_2-related health effects.

The three study areas and time periods simulated are described in section 3.1 below. Air quality modeling is used to develop the spatially varying distributions of 1-hour concentrations, as described in section 3.2. The definition of the extent and scale of the exposure modeling domain and associated air quality receptor grid is described in section 3.3. The next step in the approach is the development of an air quality scenario for each study area that reflects conditions that just meet the current standard. This step involves adjustment of the estimates resulting from the air quality modeling for each area. Section 3.4 summarizes the method used for the adjustment of the air quality concentrations to a scenario that just meets the current primary SO_2 standard. Development of the temporally varying 5-minute concentrations at each air quality receptor site is described in section 3.5.

3.1 CHARACTERIZATION OF STUDY AREAS

The study areas for this REA are Fall River, MA, Indianapolis, IN, and Tulsa, OK (Table 3-1). These study areas were selected to meet a number of individual and collective criteria. The following list includes the criteria used in considering individual study areas:

- **Design value[1] near the current standard (75 ppb).** Using recent air quality monitoring data (2011-2015), design values ranging from 50 ppb to 100 ppb were considered preferable in order to minimize the magnitude of the adjustment needed to generate air quality just meeting the current standard, therefore potentially minimizing the uncertainties in estimates of exposures associated with the adjustment approach. In considering areas with regard to this criterion, consecutive 3-year periods as far back as 2011-2013 were considered.
- **One or more air quality monitors reporting 5-minute SO_2 data for the 3-year study period.** In judging whether monitors provided such a 3-year record, completeness requirements (summarized in section 3.5) were applied for all three years to ensure the availability of adequate data for informing the ambient air concentrations used for exposure modeling. Study areas having continuous 5-minute data were preferable to those with only hourly maximum 5-minute data. There are no monitoring requirements to report continuous 5-minute data at all ambient air monitors, therefore we used this as an additional consideration after an initial screen for the top candidate areas.
- **Availability of existing air quality modeling datasets**. There are many areas in the U.S. that have chosen to model air quality for regulatory purposes, i.e., in designating areas with regard to determining the attainment of the current standard. This criterion was considered important for efficiency purposes and to maintain consistency between our assessment approach and state-level modeling regarding the years selected, sources included, emission levels and profiles, and assumptions used to predict ambient air concentrations.
- **Population size greater than 100,000.** Candidate study areas having the larger populations were given priority to provide a more robust and improved representation of exposures and risk to key at-risk populations.
- **Significant and diverse emissions sources.** Preference was given to study areas with a diverse source mix, including EGUs, petroleum refineries, and secondary lead smelting (generally reflects battery recycling). A diverse source mix allows for capturing

[1] A design value (DV) is a statistic that describes the air quality status of a given area relative to a particular NAAQS. A design value summarizes the concentrations of a criteria pollutant in terms of the statistical form of the standard for that pollutant, thus indicating whether the area meets or exceeds the standard. Consistent with the form of the SO_2 standard, SO_2 design values are calculated as the 3-year average of the annual 99^{th} percentile of the daily maximum 1-hour average concentrations (see 40 CFR 50.17). By regulation, design values calculated from monitoring data are considered to be valid if they meet specified completeness criteria, which for SO_2 are data for at least 75 percent of the sampling days in all four quarters of all three years of the period (see Appendix T to Part 50).

exposures to both large sources (e.g., emissions of 10,000-20,000 tons per year [tpy])[2] and small sources (e.g., emissions of hundreds of tpy) distributed about a study area.

In consideration of the above criteria, Fall River, MA, Indianapolis, IN, and Tulsa, OK were selected.[3] In identifying this set of study areas, we also concluded it to be desirable for the study areas, as a set, to represent different geographical regions of the U.S. The three study areas – in Massachusetts, Indiana and Oklahoma –are in three different climate regions of the U.S.: the Northeast, Ohio River Valley (Central), and South (Karl and Koss, 1984). These regions, particularly the Ohio River Valley, generally have a higher concentration of EGU and non-EGU sources of SO_2 emissions than other areas of the country (ISA, Figure 2-3). Given the objective of assessing air quality conditions that just meet the current standard, our focus, as indicated by the first criterion above, is not on the areas in the U.S. with ambient air concentrations substantially above the standard.[4] Additionally, we minimized inclusion of study areas near the ocean or large water bodies, such as the Great Lakes, given the potential for unusual atmospheric chemistry and associated transformation of SO_2 in those areas and our limited ability to accurately model such events.

We considered more than one hundred areas and multiple time periods as study area candidates. Closer examination of candidate areas and time periods led us to select the three study areas identified above and the study period of 2011 to 2013, as they best fit the above selection criteria.[5] The study areas and time periods selected – Fall River, MA, Indianapolis, IN, and Tulsa, OK (Table 3-1) – together represent an array of differing exposure circumstances for 5-minute peak SO_2 concentrations in ambient air. This array expands on the more limited set of study areas, focused in a single region of the U.S., that was addressed in the 2009 SO_2 REA. As described in subsequent sections, information for the 2011-2013 period in the three study areas was used to develop the air quality scenarios that represent conditions just meeting the current

[2] While there may be other sources having similar or greater SO_2 emissions, design values for the ambient monitors surrounding these other sources may not necessarily fall within that particular selection criterion. Again, having monitor design values at or near the existing standard is considered important in limiting the magnitude of uncertainty associated with adjusting concentrations that just meet the existing standard.

[3] Further investigation of available information for potential study area locations with regard to the criteria identified above resulted in the identification of the selected three study areas, for two of which existing air quality modeling datasets were available. Such datasets were not available for many of the potential study areas referred to as candidates in the REA Planning Document (e.g., Detroit and Savannah).

[4] This objective of the REA and, more specifically, the design value criterion used to identify candidate study areas for the REA differs from the criteria used in selecting the six focus areas in the ISA. The selection criteria used to identify focus areas in the ISA did not consider ambient monitoring concentration levels, and as such, four of the six ISA focus areas would not meet the above REA design value criterion alone (ISA, section 2.5.2.2).

[5] Use of this time period (2011-2013) in these three study areas, in which concentrations were closer to the current standard than indicated in more recent data, allowed us to apply a smaller adjustment in developing the air quality scenarios for just meeting the current standard, thus reducing any associated uncertainty.

standard for which this REA has estimated exposures and risks to at-risk populations from SO_2 concentrations in ambient air.

Table 3-1. General features of the study areas selected for the exposure and risk assessment.

Study Area	Geographic Region	# of Monitors in Exposure Modeling Domain [a] Reporting 5-Minute Data (# with Continuous Data)	2011-2013 DV [b] (ppb)	Population in Exposure Modeling Domain [a c]	# of Sources emitting >100 tons per year [d] in Exposure Modeling Domain	Source Types [e]
Fall River, MA	New England	1 (1)	64	183,874	1	EGU
Indianapolis, IN	Ohio River Valley	3 (3)	78	547,968	4	EGUs, secondary lead smelter, airport
Tulsa, OK	South	4 (4)	55	257,423	3	EGU, petroleum refineries

[a] Delineation of the exposure modeling domain is described in section 3.4; it includes the area within 10 km of the sources with SO_2 emissions above 100 tons in 2011, 2012 or 2013 and inclusive of the monitors with 5-minute data.
[b] Highest monitor-based design value in exposure modeling domain.
[c] Population sizes are drawn from 2010 U.S. Census.
[d] This reflects information in 2011 National Emissions Inventory. As described in section 3.2, other sources are also reflected in the air quality modeling, either explicitly or via the addition of study-area-specific concentrations.
[e] This reflects sources counted in column to the left of this one. As described in section 3.2, other sources are also reflected in the air quality modeling, either explicitly or via the addition of study-area-specific concentrations.

3.2 AIR QUALITY MODELING

The EPA's preferred model for near-field dispersion, AERMOD (U.S. EPA, 2016a, b), was used to generate 1-hour concentrations for the 3-year period, 2011-2013, across the exposure modeling domains for the three study areas: Fall River, MA, Indianapolis, IN, and Tulsa, OK. In addressing the development of model inputs and specifications, as well as performing the modeling runs themselves, the steps listed below were performed for all three study area modeling domains.

(1) **Collected and analyzed general input parameters.** Meteorological data, processing methodologies used to derive input meteorological fields (e.g., temperature, wind speed, precipitation), and information on surface characteristics and land use were needed to help determine pollutant dispersion characteristics, atmospheric stability and mixing heights (section 3.2.1).

(2) **Defined sources and estimated emissions.** The modeled emission sources included major stationary emission sources within the domain (section 3.2.2).

(3) **Defined air quality receptor locations.** Receptor locations were identified for the dispersion modeling at varying spatial scale (depending on distance from source to receptor) from 2 km to 100 m (section 3.2.3).

(4) **Calculated background concentrations.** In this context the phrase "background concentrations" refers to SO_2 concentrations resulting from sources (nearby and distant) other than those whose emissions are explicitly modeled. These concentrations were calculated based on ambient air monitoring data that exclude hours of the day that were most likely influenced by the modeled emission sources (section 3.2.4).

(5) **Estimated concentrations at receptors**. Full annual time series of hourly concentration were estimated for 2011-2013 by summing concentration contributions from each of the emission sources at each of the defined air quality receptors (section 3.2.5).

Details regarding both modeling approaches and input data used are provided below with supplemental information regarding model inputs and methodology provided in Appendices A, B, and C. To ensure use of the appropriate local data for the simulated time periods, as well as efficiency and consistency for these areas, we drew on information for the Indianapolis and Tulsa study areas (e.g., stack locations, building parameters, etc.) that had been developed for regulatory purposes.[6,7] Information for the Fall River study area was developed specifically for this assessment in a manner that was technically appropriate and generally consistent with that for the other two areas. The sections below summarize development of the information described in the steps listed above for all study areas. Figures 3-1 to 3-3 show the locations of the upper and surface meteorological stations, the modeled SO_2 emission sources, and the ambient monitoring sites used for predicting air quality used in this REA. Because some of the meteorological stations and emissions sources were located outside of the general study area,[8] two maps are provided for each study area: one map encompassing all of the features and the second map focused on those features closest to or within each study area.

[6] For the Indianapolis study area, we drew on the modeling performed by Indiana Department of Environmental Management for Indiana's State Implementation Plan (SIP) for the Marion County SO_2 nonattainment area. This documentation is available at: http://www.in.gov/idem/airquality/files/attainment_so2_multi_2015_demo_attach_k.pdf.

[7] For the Tulsa study area, we drew on the modeling performed by Oklahoma Department of Environmental Quality to address regulatory Prevention of Significant Deterioration (PSD) requirements for refineries in the Tulsa area. This information is available for Permits 2012-1062-TVR2 M-9 and 2010-599-TVR M-7 at: http://www.deq.state.ok.us/aqdnew/permitting/PermitsIssuedDuringPastYear.html.

[8] For better visualization of the meteorological stations, emission sources, and the ambient monitors used to estimate air quality for this assessment, the area highlighted is an approximation based on census tracts that encompass the actual exposure study area (section 3.3) which is comprised of a subset of census blocks within those same census tracts.

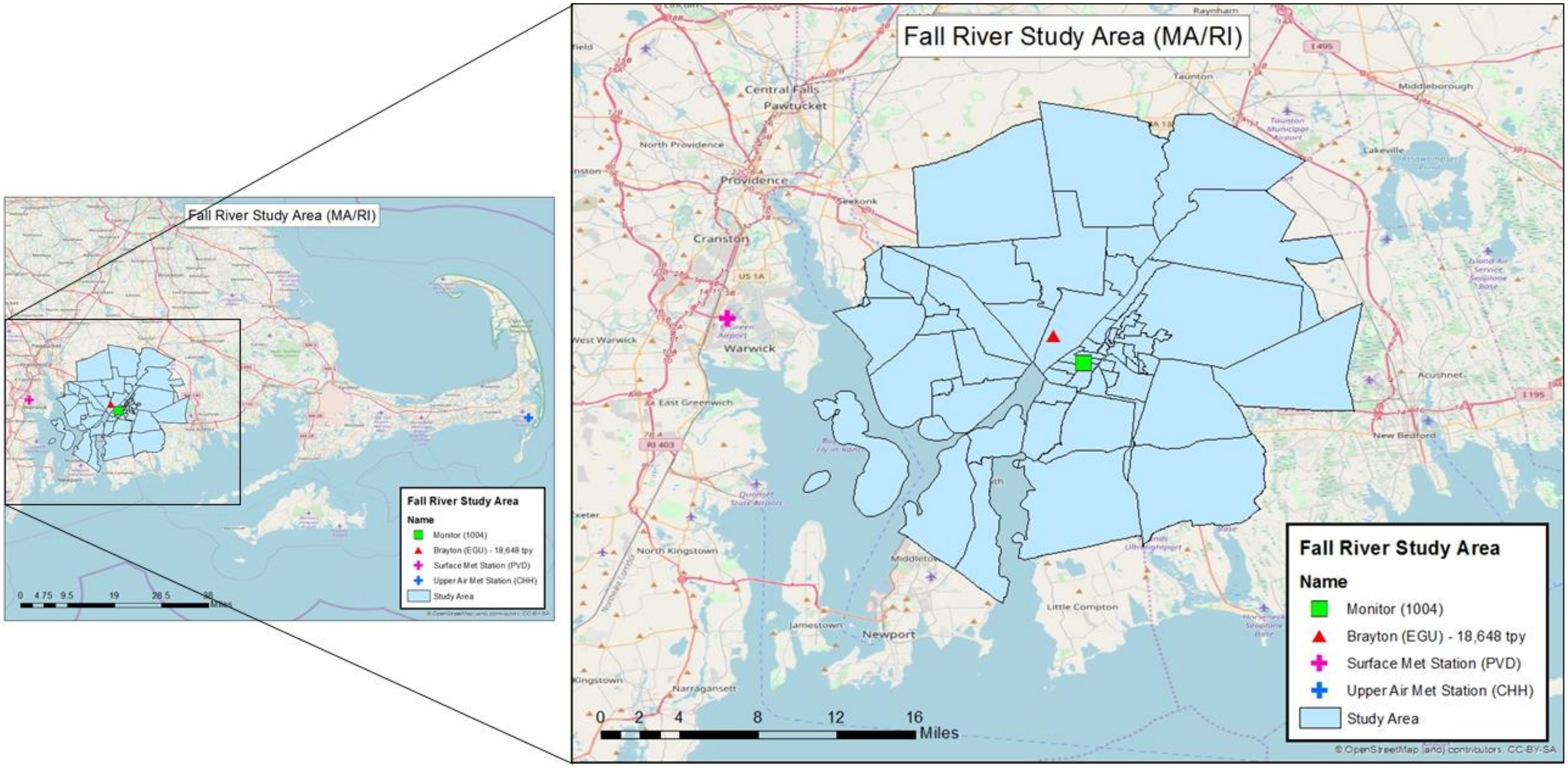


Figure 3-1. Location of surface and upper air meteorological stations, SO_2 emissions sources, and ambient monitors used to predict ambient air quality in the Fall River study area. Also included is source type and 2011 NEI SO_2 emissions.

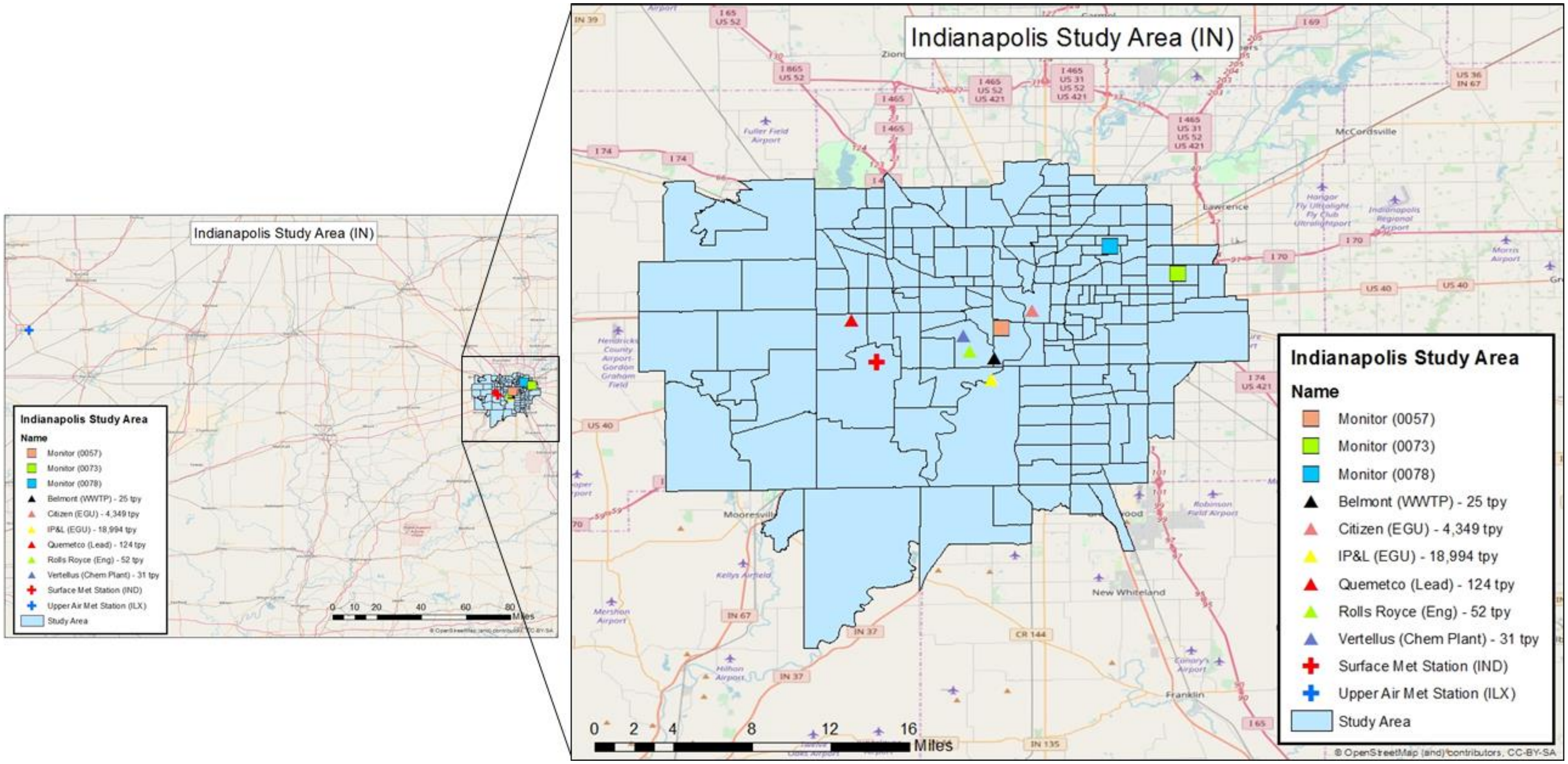


Figure 3-2. Location of surface and upper air meteorological stations, SO_2 emissions sources, and ambient monitors used to predict ambient air quality in the Indianapolis study area. Also included is source type and 2011 NEI SO_2 emissions.

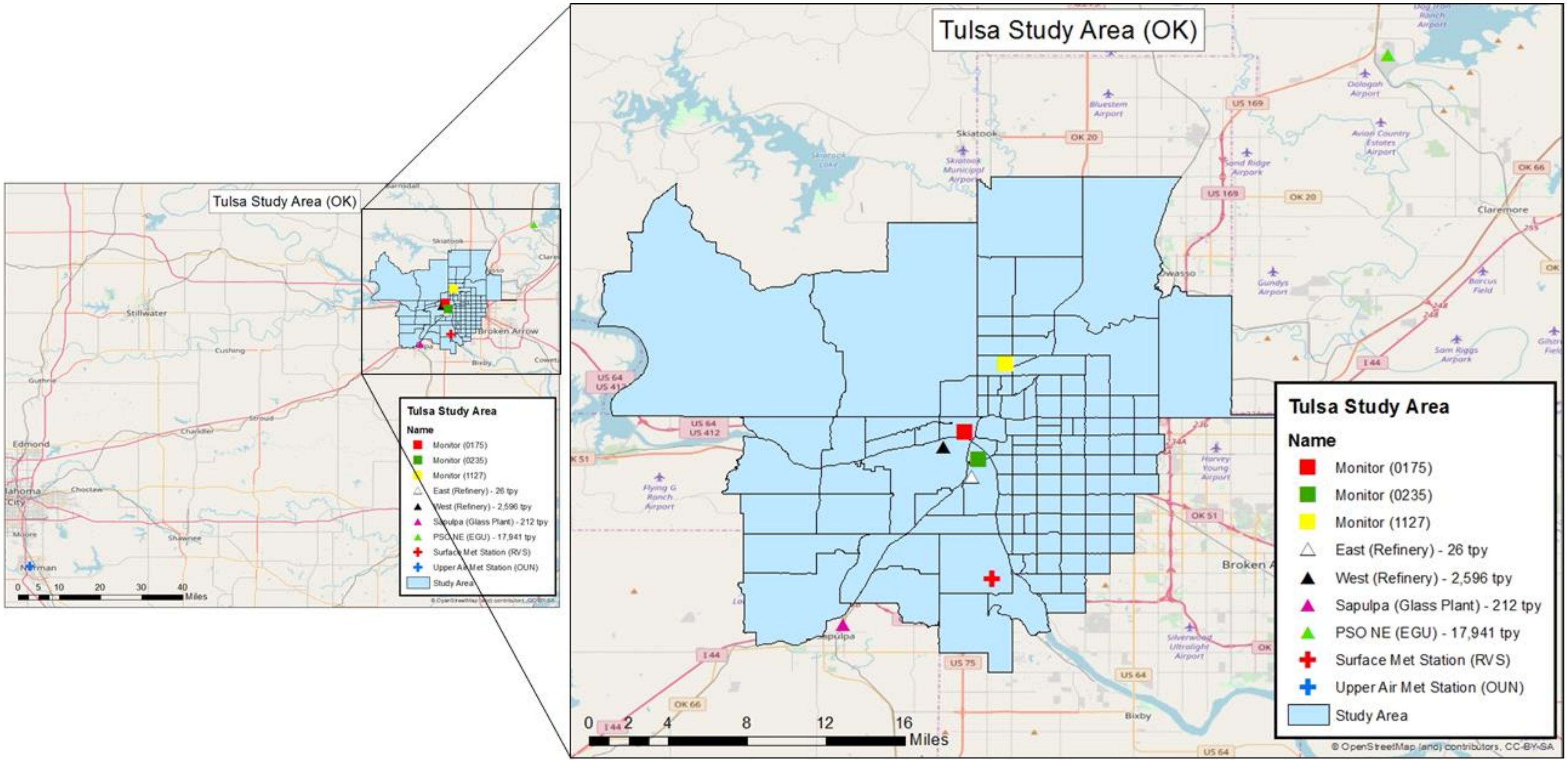


Figure 3-3. Location of surface and upper air meteorological stations, SO_2 emissions sources, and ambient monitors used to predict ambient air quality in the Tulsa study area. Also included is source type and 2011 NEI SO_2 emissions.

3.2.1 General Model Inputs

3.2.1.1 Meteorological Inputs

All meteorological data used for the AERMOD dispersion model simulations were processed with the AERMET meteorological preprocessor, version 16216 (U.S. EPA, 2016c) using regulatory options. The National Weather Service (NWS) served as the source of input meteorological data for AERMOD. Tables 3-2 and 3-3 list the surface and upper air NWS stations chosen for the three study areas. The NWS hourly surface data are archived in the Integrated Surface Hourly (ISH) database for which there is a potential concern for a high incidence of calms and variable wind conditions. This is due to how the hourly data are reported from the Automated Surface Observing Stations (ASOS) in use at most NWS stations. Wind speeds less than three knots are assigned a value of zero knots, and the definition used for a variable wind observation (wind direction that varies more than 60° in a 2-minute observation) may include wind speeds up to 6 knots, but with a wind direction that is reported as missing. The AERMOD model currently cannot simulate dispersion under these conditions. This issue was addressed by reducing the number of calms and missing winds in the surface data for each of the three NWS surface stations using separately archived 1-minute averaged wind data from the ASOS stations. Low wind speeds and wind direction are retained in the 1-minute ASOS data. Hourly average wind speeds and directions were calculated using the 1-minute wind data to supplement the hourly wind data in the ISH format. The 1-minute data were processed with AERMINUTE, version 15272 (U.S. EPA, 2015a). AERMINUTE performs quality assurance procedures on the 1-minute data files, computes the hourly averages of wind speed and direction, and outputs the hourly averages in a data file that can be directly input into AERMET.

Table 3-2. National Weather Service surface stations for meteorological input data in study areas.

Study Area	Station	Identifier	WMO (WBAN)	Latitude (degrees)	Longitude (degrees)	Elevation (m)	GMT Offset (hours)
Fall River, MA	Providence	PVD	725070 (14765)	41.7225	-71.4325	19	-5
Indianapolis, IN	Indianapolis International Airport	IND	724380 (93819)	39.725170	-86.281680	241	-5
Tulsa, OK	Tulsa R L Jones Jr Airport	RVS	723564 (53908)	36.042441	-95.990166	192	-6

Table 3-3. National Weather Service upper air stations for meteorological input data in study areas.

Study Area	Station	Identifier	WMO (WBAN)	Latitude (degrees)	Longitude (degrees)	Elevation (m)	GMT Offset (hours)
Fall River, MA	Chatham, MA	CHH	744940 (14684)	41.67	-69.97	12	-5
Indianapolis, IN	Lincoln, IL	ILX	745600 (04833)	40.15	-89.33	178	-6
Tulsa, OK	Norman, OK	OUN	723560 (13968)	35.23	-97.47	354	-6

3.2.1.2 Surface Characteristics and Land Use Analysis

The AERSURFACE tool, version 13016 (U.S. EPA, 2013) was used to determine surface characteristics (e.g., albedo, Bowen ratio, and surface roughness) for input to AERMET. Surface characteristics were calculated for the location of the ASOS meteorological towers, which were approximated by using aerial photos and the station history from the National Centers for Environmental Information (NCEI). AERSURFACE utilizes 1992 land cover data from the National Land Cover Dataset (NLCD). Land cover data was obtained from the Multi-Resolution Land Characteristics (MRLC) consortium website.[9] Each of the three surface meteorological stations are located at an airport and were specified accordingly in AERSURFACE.

Though the current version of AERSURFACE is limited to processing older land cover data for input to AERMET, a review of historical and more recent satellite imagery indicates there have not been substantial changes in the land cover within the area immediately

[9] https://www.mrlc.gov

surrounding the meteorological towers for the three modeled sites between 1992 and 2011.[10] For each of the three sites that were modeled, the surface meteorological observations were collected at NWS stations located at airports. The meteorological towers at these airports are located in grassy areas near or between runways. Surface roughness is derived as an inverse-distance weighted average of the land cover within a 1.0 km radius centered on the meteorological tower. Thus, the land cover that is nearest to the tower, where there is the least amount of change over time, has the greatest influence on the derived roughness value. Bowen ratio and albedo, on the other hand, are derived from a 10 km × 10 km area centered on the meteorological tower. Bowen ratio and albedo represent an average of the land cover across the 10 km × 10 km area in which each land cover pixel is weighted equally. Isolated areas where the land cover has changed substantially over time have little effect on the average value of Bowen ratio and albedo within the 10 km × 10 km area.

AERSURFACE allows for the surface roughness length to be defined by up to 12 wind sectors with a minimum arc of 30 degrees each. For each of the three ASOS stations, roughness was estimated for each of 12 sectors, beginning at 0 degrees through 360 degrees (i.e., 0-30, 30-60, 60-90, etc.). The wind sectors for each of the three surface stations are illustrated in Appendix A. The AERSURFACE default month-to-season assignments were used for Tulsa, and reassignments were performed for both Indianapolis and Fall River. The monthly seasonal assignments input to AERSURFACE for each of the three surface stations are shown in Table 3-4. Surface characteristics were output by month. Note that there are two winter options: 1) winter with no snow (or without continuous snow) on the ground the entire month and 2) winter with continuous snow on ground the entire month.[11] A month was considered to have continuous snow cover if a snow depth of one inch or more was reported for at least 75% of the days in the month.

Table 3-4. Monthly seasonal assignments input to AERSURFACE.

Study Area	Winter (continuous snow)	Winter (no snow)	Spring	Summer	Fall
Fall River, MA	-	Dec, Jan, Feb, Mar	Apr, May	Jun., Jul, Aug	Sep, Oct, Nov
Indianapolis, IN	-	Dec, Jan, Feb, Mar	Apr, May	Jun., Jul, Aug	Sep, Oct, Nov
Tulsa, OK	-	Dec, Jan, Feb	Mar, Apr, May	Jun., Jul, Aug	Sep, Oct, Nov
Seasonal definitions: Winter - Late fall after frost and harvest, or winter with no snow; Spring - Transitional spring with partial green coverage or short annuals; Summer - Midsummer with lush vegetation; Fall - Fall with unharvested cropland					

[10] Google Earth was used to evaluate the land use/land cover in the area immediately surrounding the meteorological tower, out to a distance of 1 km, and similarly in the region around the airport, out to a distance of 5 km.

[11] For many of the land cover categories in the 1992 NLCD classification scheme, the designation of winter with continuous snow on the ground would tend to increase wintertime albedo (reflectivity) and decrease wintertime Bowen ratio (sensible to latent heat flux) and surface roughness compared to the winter with no snow or without continuous snow designation.

AERSURFACE also requires information about the climate and surface moisture at the surface station. The station has to be categorized as either arid or non-arid. Each of the three surface stations were categorized as non-arid in AERSURFACE. Surface moisture is based on precipitation amounts and is categorized as either wet, average, or dry. For the three surface stations, 2010 local climatological data from the NCEI was used to look at 30 years (1981-2010) of monthly precipitation. The 30^{th} and 70^{th} percentiles of precipitation amounts were calculated separately for each of 12 months (January through December) based on the 30-year period. The precipitation amount for each month in 2011-2013 was then compared to the 30^{th} and 70^{th} percentiles for the corresponding month. Months during which precipitation was greater than the 70^{th} percentile were considered wet, while months that were less than the 30^{th} percentile were considered dry. Months within the 30^{th} and 70^{th} percentile range were considered average. AERSURFACE was run for each moisture condition to obtain monthly values for wet, dry, and average conditions. Using the AERSURFACE output for each of the three moisture categories, a separate set of monthly surface characteristics was compiled for each of the three years for input to AERMET. The monthly categorization of the surface moisture at each of the locations is shown in Table 3-5. Refer to Appendix A for a complete listing of the surface characteristic values input to AERMET for each surface station and a detailed discussion of the meteorological data preparation.

Table 3-5. Monthly surface moisture categorizations for the three study areas.

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Fall River, MA												
2011	Avg.	Wet	Dry	Wet	Avg.	Wet	Wet	Wet	Wet	Wet	Wet	Avg.
2012	Avg.	Dry	Dry	Avg.	Wet	Wet	Avg.	Wet	Wet	Wet	Dry	Wet
2013	Dry	Wet	Dry	Dry	Avg.	Wet	Avg.	Wet	Wet	Dry	Wet	Wet
Indianapolis, IN												
2011	Wet	Wet	Wet	Wet	Wet	Wet	Dry	Dry	Wet	Wet	Wet	Wet
2012	Wet	Avg.	Wet	Avg.	Dry	Dry	Dry	Wet	Wet	Wet	Dry	Avg.
2013	Wet	Wet	Wet	Wet	Wet	Wet	Dry	Dry	Wet	Wet	Wet	Wet
Tulsa, OK *(Moisture conditions at RVS are based on precipitation data from Tulsa International Airport, TUL)*												
2011	Dry	Wet	Dry	Wet	Dry	Dry	Dry	Wet	Dry	Dry	Wet	Avg.
2012	Dry	Avg.	Wet	Avg.	Dry	Wet	Dry	Wet	Dry	Avg.	Dry	Dry
2013	Wet	Wet	Dry	Avg.	Avg.	Dry	Wet	Wet	Dry	Wet	Avg.	Avg.
Moisture categories were defined by comparing existing year/month precipitation values with 30-year monthly precipitation data set: Wet (>70^{th} percentile); Dry (<30^{th} percentile); Avg. (within 30^{th} and 70^{th} percentile)												

3.2.2 Stationary Sources Emissions Preparation

3.2.2.1 Emitting Sources and Locations

The modeling approach in all three study areas involved modeling key sources as point sources and accounting for other sources through the use of additional study-area-specific concentrations (see section 3.2.4). The facilities modeled as point sources included all those emitting more than 100 tpy of SO_2 in 2011, as well as some in Indianapolis that were somewhat smaller (Table 3-6). These facilities were selected from version 2 of the 2011 National Emissions Inventory (NEI)[12] and paired to a representative surface meteorological station. Any stacks listed as in the same location with identical temporal profiles and identical release parameters within a certain tolerance (typically to the nearest integer value) were aggregated into a single stack to simplify modeling, but all emissions were retained. For facilities with an SO_2 emission total exceeding 1,000 tpy in 2011, every stack emitting more than 1 tpy was included in the modeling inventory.

Table 3-6. Facilities with point sources included in the air quality modeling domain for each study area.

Study Area	Facility Name	NEI ID
Fall River, MA [a]	Brayton Point Energy (EGU)	5058411
Indianapolis, IN	Belmont Advanced Wastewater Treatment Plant [b]	4885211
	Citizens Thermal, formerly Indianapolis Power and Light	4885311
	IPL – Harding Street Generating Station	7255211
	Rolls Royce Corporation (combustion engine manufacture and testing) [b]	7972011
	Vertellus Specialties, formerly Reilly Industries and Reilly Tar and Chemical (chemical manufacturing)	7972111
	Quemetco (lead battery recycling facility)	8235411
Tulsa, OK [c]	Public Service Co. of Oklahoma (PSO) Northeastern Power Station	8212411
	Sapulpa Glass Plant	7320611
	Tulsa Refinery West	8402711
	Tulsa Refinery East	8003911

[a] Contributions to ambient concentrations from another facility emitting more than 100 tpy (SEMASS Partnership municipal waste combustor [8127611]), although 30 km away, are accounted for by the additional study-area-specific concentrations for Fall River (see section 3.2.4).
[b] These sources, although having 2011 NEI emissions under 100 tons, were included based on proximity to nearby monitoring locations and previous modeling for Indianapolis and Tulsa.
[c] There are facilities in the region outside of the immediate study area emitting more than 100 tpy (e.g., Oklahoma Gas & Electric Company Muskogee Generating Station [8506011], however, they are outside the nominal distance (50 km) used for dispersion modeling. Note also, contributions to ambient concentrations from any emission sources not explicitly modeled and potentially influencing ambient concentrations in the study area are accounted for by the additional study-area-specific concentrations for Tulsa (see section 3.2.4).

[12] See: https://www.epa.gov/air-emissions-inventories/2011-national-emissions-inventory-nei-technical-support-document

The locations of all emitting stacks that were modeled were corrected based on GIS analysis or by using locations identified in the local information developed by the state of Indiana for modeling for Indianapolis and the state of Oklahoma for Tulsa.[13] This was necessary because many stacks in the NEI are assigned the same location, which often corresponds to a location in the facility rather than the actual stack locations. NEI sources were mapped to AERMOD sources based on matching stack parameters and temporal profiles within the same facility. The release heights and other stack parameters were taken from the values listed in the 2011 NEI. Table B-3-1 (in Appendix B) lists all stacks in all domains.

3.2.2.2 Source Terrain Characterization

With the exception of sources at Quemetco and fugitive sources at Rolls Royce in Indianapolis, all source elevations for the three study areas were calculated in AERMAP, version 11103 (U.S. EPA, 2016d). Source elevations at Quemetco and fugitive sources at Rolls Royce were determined by ArcGIS overlays of the sources and National Elevation Data (NED).

3.2.2.3 Emissions Data Sources

Data for the parameterization of major facility point sources in the modeling domains comes primarily from these sources: the 2011 NEI (U.S. EPA, 2015b),[14] point source submissions to the NEI database for the years 2012 and 2013,[15] the Air Markets Program data (CAMD database) (U.S. EPA, 2017a), and temporal emission profile information from the EPA's 2011v6.3 Emissions Modeling Platform (U.S. EPA, 2016e). The NEI database contains stack locations, emissions release parameters (i.e., height, diameter, exit temperature, exit velocity), and annual SO_2 emissions. The CAMD database has information on hourly SO_2 emission rates for all the electric generating units (EGUs) in the U.S. where the units are boilers or equivalent, each of which can have multiple stacks. For sources that did not have hourly data in the CAMD database, annual total emissions data from the NEI were converted into the hourly temporal profiles required for AERMOD according to temporal profiles that are part of the EPA's 2011v6.3 emissions modeling platform.

---

[13] As noted in section 3.2 above, local information was provided by these states in documentation developed for SIP and PSD-related purposes.

[14] We consider the 2011 NEI is the most appropriate emissions data set to use for modeling the 3-years of air quality in this REA because the exposure period used is based on 2011-2013 ambient monitor data (and the associated meteorology).

[15] Annual total emissions for the largest point sources are reported to the NEI each year by the State air agencies. Every third year (e.g., 2011, 2014), emissions for all point sources are to be reported to the NEI by the State air agencies. Submissions to the NEI may also include any needed changes to the facility information for point sources (e.g., locations, stack parameters, control devices), as this information is stored persistently in the NEI database between NEI submission cycles and is updated as needed.

The emissions information needed for running AERMOD was drawn from this array of information sources (detailed information is provided in Appendix B). For EGU sources, the more detailed information (e.g., hourly emissions values) were drawn from the CAMD database and annual estimates from the NEI. For sources other than EGUs for which hourly SO_2 emissions estimates were not available in the CAMD database, temporal profiles were used to prepare the hourly emissions factors as described in Appendix B.

The designation of sources in the three study areas as urban or rural reflected information about the source and surrounding area. The urban/rural designation of a source is important in determining the boundary layer characteristics that affect the model's prediction of downwind concentrations. It is particularly important for SO_2 modeling because AERMOD invokes a 4-hour half-life for urban SO_2 sources (U.S. EPA, 2016a, section 7.2.1.1) to account for SO_2 removal by conversion to sulfuric acid (catalytic and photochemical) and adsorption on to particulate matter (Turner, 1964).[16] For Fall River, a rural designation was used based on land use data, the fact that the stacks at Brayton were tall, and the AERMOD Implementation Guide (U.S. EPA, 2016g) recommendation to use a rural designation when modeling tall stacks in urban areas. Classifying tall stacks with buoyant releases as urban sources in urban areas may artificially limit plume height, thus artificially increasing modeled ground level concentrations. The use of the AERMOD urban option for these sources may not be appropriate given that the actual plume is likely to be transported over the urban boundary layer. For Indianapolis, all sources were classified as urban sources based on having a broadly defined urban population of 1,000,000, consistent with the classification in the SIP modeling. For Tulsa, all sources were classified as urban based on having a broadly defined urban population of 396,466, consistent with the classification in the PSD modeling.

Building downwash parameters for Indianapolis and Tulsa were set based on local information available from Indiana and Tulsa state modeling work. Given the lack of building information available in Fall River, building downwash was not used in modeling for this study area.

3.2.3 Air Quality Receptor Locations

Among the three study areas, the sizes of the air quality modeling domain and receptor grid varied in consideration of differences such as number, size, and distribution of the key emissions sources. The domains and receptor grids for Indianapolis and Tulsa drew on the approach used by Indiana and Oklahoma in modeling these areas for their SIP and for PSD

[16] For urban sources, AERMOD accounts for the urban heat island effect on increasing mixing heights for hours under atmospheric stable conditions. Details on determining the urban or rural status of sources can be found in U.S. EPA (2016a), U.S. EPA (2016f), and U.S. EPA (2016g).

purposes. Where these domains were larger than the areas of interest for the exposure assessments, the receptor grids were subset to receptors that encompassed the census blocks of interest for the exposure assessment as described in section 3.3 below. The full air quality modeling domain for Indianapolis was 38 km × 32 km with receptor spacing ranging from 2 km at the edges, to 1 km, 500 m, 250 m, and 100 m near the emission sources, with fence line receptors included.[17] The Tulsa domain was 26 km × 29 km and receptor spacing ranged from 1 km at the edges to 666.67 m, 250 m, and 100 m near the sources, with fence line receptors also included. For Fall River, we generated a domain (20 km × 20 km receptor grid with 500 m spacing) that specifically fit the needs of the exposure assessment. Receptor elevations and hill heights for all three areas were obtained from AERMAP.

3.2.4 Concentrations Associated with Sources Not Explicitly Modeled

Concentrations associated with sources of SO_2 that were not explicitly modeled in all three study areas (e.g., source emissions from outside the modeling domain in addition to emissions from sources within the domain that were not explicitly modeled with AERMOD) were separately estimated and added to the AERMOD modeled concentrations to produce the hourly concentrations at each receptor. For example, for Fall River these concentrations were approximated to account for the impacts from SEMASS Partnership given its distance (~30 km) from the Fall River emission source of interest (Brayton EGU), rather than including SEMASS Partnership as a point source in the AERMOD modeling run.

For all three study areas, these concentrations were calculated in terms of three-year averages of seasonal-hour-of-day concentrations.[18] This approach generally relied on the use of ambient air monitoring data from a designated monitor (i.e., one not receiving direct impact from emission sources modeled in the domain). Measurements from this monitor were excluded, as recommended in the EPA air quality modeling guidance (U.S. EPA, 2016a, f), during times when the sources that were explicitly modeled were potentially impacting monitor concentrations and were informed by monitor siting relative to the modeled sources and wind direction.[19] For Fall River, monitor 250051004 (see Figure 3-1) was used for this purpose. Hours

[17] The air quality modeling receptor grids utilized varying spatial resolution within the grids, as is customary in most regulatory modeling applications. The exact placement of receptors usually depends on individual state modeling guidance for dispersion modeling for regulatory applications. This accounts for the varying range of receptor grids in the assessment for Indianapolis and Tulsa. Receptors are normally placed in locations of ambient air, i.e. where the general public has access and along fencelines of the modeled sources. Receptors are usually spaced close together near the modeled sources to capture concentration gradients near the sources, and they are spaced with decreasing spatial resolution farther away from the sources.

[18] This approach was implemented as recommended in the EPA's modeling guidance for SO_2 (U.S. EPA 2016f).

[19] Wind direction data was obtained from the surface meteorological stations representing each study area.

when wind direction was from the west to north (270° to 360°) were excluded from the calculation to remove the impacts from the source that was explicitly modeled (Brayton EGU). For Indianapolis, monitor 180970078 (northern monitor; see Figure 3-2) was used. Hours with wind directions between 170° and 270° were excluded to eliminate impacts from the modeled sources in that study area. For Tulsa, monitor 401431127 (located north of the refineries, see Figure 3-3) was used. Hours when the wind direction was either 90° to 140° or 270° to 6° were excluded to eliminate impacts from the two refineries or the PSO Northeastern power station. Table 3-7 shows the seasonal-hour-of-day concentrations estimated to result from source emissions not explicitly modeled in the AERMOD runs for the three study areas.[20]

[20] Use of this approach to estimate concentrations associated with source emissions not modeled contributes to uncertainty in the exposure and risk estimates and is summarized in Table 6-3 below.

Table 3-7. SO_2 concentrations (ppb) used to account for source emissions not explicitly modeled in the three study areas, stratified by season and hour of day.

Hour	Fall River				Indianapolis				Tulsa			
	Winter	Spring	Summer	Fall	Winter	Spring	Summer	Fall	Winter	Spring	Summer	Fall
1	4.07	5.47	9.07	9.43	10.93	6.73	5.03	3.67	2.27	1.27	5.50	1.20
2	5.27	8.43	6.37	7.07	10.60	6.63	3.93	4.73	2.33	0.87	2.60	1.50
3	4.77	4.70	9.13	9.13	10.37	6.40	4.20	4.33	1.83	0.40	4.30	0.93
4	7.30	5.40	7.63	12.23	8.87	6.73	4.57	3.43	1.83	0.50	0.70	1.47
5	8.03	4.80	7.40	10.37	8.83	6.87	7.63	4.70	2.03	1.37	0.60	1.70
6	6.23	4.97	8.00	11.03	11.67	5.23	3.83	5.33	1.93	0.47	8.30	1.43
7	9.30	6.83	7.83	11.27	13.40	5.37	5.20	5.40	1.57	1.03	0.80	1.47
8	8.27	6.07	7.47	8.33	10.00	6.07	5.93	6.47	2.33	3.90	1.20	2.63
9	7.17	5.80	7.30	8.20	7.77	7.50	30.7	7.10	1.93	1.23	1.33	1.50
10	8.13	5.43	7.27	9.40	13.07	9.73	25.73	15.13	2.90	2.37	0.93	1.43
11	8.57	9.30	10.50	7.47	10.20	9.07	23.27	40.57	2.80	1.87	1.53	2.63
12	8.43	7.80	18.37	8.90	12.70	8.63	17.63	37.93	5.30	2.17	2.20	2.67
13	8.77	11.83	15.90	7.50	17.63	5.93	14.83	21.83	6.13	2.30	2.40	5.23
14	9.27	8.33	16.93	7.00	13.13	5.60	9.50	11.07	2.80	2.30	3.03	2.90
15	8.00	3.30	6.40	4.00	13.13	16.33	7.40	7.97	1.80	1.67	2.00	2.20
16	6.83	2.33	6.00	3.67	7.53	4.87	9.90	12.53	3.10	1.97	2.47	2.83
17	8.93	3.60	4.33	3.03	6.97	6.30	8.53	25.10	3.30	3.60	2.13	4.17
18	5.80	2.47	3.63	2.70	11.27	11.37	9.97	16.33	4.27	3.67	5.77	4.00
19	4.43	2.30	3.27	2.87	6.77	9.10	7.47	10.83	2.87	1.47	1.50	2.20
20	4.33	2.03	3.20	2.73	9.57	4.93	10.20	7.60	2.33	2.87	1.83	2.53
21	4.07	2.30	3.13	2.67	10.57	5.87	6.13	6.57	2.57	2.67	1.33	2.00
22	3.63	2.10	2.97	2.57	12.17	4.27	10.30	5.47	2.63	1.37	0.93	2.20
23	3.70	2.60	3.07	2.60	6.13	6.13	10.73	4.00	3.67	1.03	0.67	2.30
24	4.80	2.80	6.77	7.93	5.67	6.27	6.63	3.9	3.17	1.43	2.17	1.87

3.2.5 Hourly Concentrations at Air Quality Model Receptors

Once all model inputs have been created, i.e. hourly meteorology, emissions, building parameters, etc., the AERMOD dispersion model is run to estimate hourly concentrations for each study area. AERMOD reads the hourly meteorological data files, pairs the hourly meteorology with the appropriate emissions and building parameters for each hour and uses Gaussian plume theory to calculate an hourly concentration at each receptor. AERMOD then outputs the hourly concentrations to a file that can be used in the exposure assessment.[21] An initial evaluation of the modeled concentrations based on comparison to the full distribution of monitored concentrations can be found in Appendix D.[22] Briefly, modeled concentrations were compared to ambient air measurements using two approaches: calculated design values and simple Q-Q (quantile-quantile) plots of the 1-hour, 3-hour, and 24-hour average concentrations. Overall, for the three modeled areas, the modeled concentrations were comparable to the ambient air measurements, although there were instances of over- and under-prediction of concentrations at the upper percentiles of the concentration distribution. When evaluating on an annual basis, model-to-monitor agreement tended to be best using the 2011 concentrations.

To augment the model-monitor evaluation of hourly SO_2 concentrations in ambient air presented in Appendix D, we performed an additional evaluation focused on air quality model estimates during time periods with relatively greater potential for population exposures.[23] The context for this air quality modeling performance evaluation[24] is particular to the intended purpose of the air quality modeling in providing estimates of 1-hour concentrations across the exposure modeling domain that are used with spatially limited monitoring data to estimate short-term exposure concentrations, especially those in outdoor microenvironments. The focus is on

[21] For this assessment, AERMOD output the hourly concentrations resulting from emissions from each of the largest sources in each study area separately. These concentrations were used to develop a factor for adjusting concentrations such that total concentrations (from all sources) just meet the current standard (section 3.4). After adjustment, modeled concentrations were combined along with the estimated concentration contribution from source emissions not explicitly modeled and were then used in estimating population exposures.

[22] In this section, “modeled concentrations” and “unadjusted model estimates” refers to the concentrations derived by adding the concentrations estimated to result from sources not explicitly modeled (section 3.2.4) to the AERMOD outputs.

[23] While the continuous time-series of hourly concentrations estimated by the air quality modeling is not expected to precisely reflect that of the monitor measurements, some consistency with regard to when relatively higher concentrations occur (e.g., daytime vs nighttime) is particularly desirable for use in exposure modeling and provides a measure of confidence with respect to the intended use of the ambient air concentration estimates and in estimating population exposures.

[24] Given the specialized use for the air quality model predictions, we recognize the importance of performance considerations that may differ from those common in evaluations of air quality modeling for regulatory purposes. For example, an area of interest in our evaluation described here is consideration of the occurrence of peak concentrations during times with greater (*versus* lesser) population exposure potential.

outdoor microenvironments (and hence, ambient air concentrations) given the overwhelming influence of these MEs on population exposure estimates (see section 5.2). We also recognize that participation in outdoor events is typically influenced by seasonal and diurnal variability in activity patterns. For example, more people spend time outdoors when the weather is comfortable (e.g., temperate spring mornings or autumn afternoons) and during daylight hours than when conditions are opposite (e.g., cold winter nights). Accordingly, this evaluation of the modeled and measured hourly concentrations considers important time-of-day and time-of-year stratifications.[25]

While the estimated exposures in this REA do not utilize the AERMOD predictions without adjustment, the evaluation summarized here (on the unadjusted model estimates and monitor measurements) is considered to provide some perspective on the uncertainty that may be contributed to the spatial and temporal pattern of hourly concentrations estimated by AERMOD that then feeds into the air quality adjustment step (described in section 3.4) and to the development of the 5-minute concentrations in each study area (described in section 3.5). Because this data evaluation was performed using the unadjusted ambient air quality, conclusions drawn from this evaluation, while informative regarding the overall model performance, are not directly transferrable to the hypothetical air quality scenario simulated for the REA main body results, *per se*. Additionally, we were not able to develop directly comparable modeling and monitoring datasets for our hypothetical air quality scenario (i.e., air quality adjusted to just meet the current standard) because the adjustment approach applied to the model estimates to create this scenario uses a proportional factor to adjust the primary source concentration contribution at each receptor, while holding all other source concentration contributions unadjusted (section 3.4). Accordingly, this approach cannot be applied to the ambient air monitoring concentrations. Thus, the evaluation provided here is simply intended to be somewhat informative, particularly with regard to considering the extent to which the relatively higher concentration events predicted by the modeling occur in the same seasons and portion of the day (daytime vs nighttime) as the relatively higher concentration monitor events.

We focus our evaluation here on the relatively higher monitor concentrations (i.e., the upper part of the concentration distribution) that occur during daytime hours in the spring, summer and fall seasons, and using the monitors having the highest design value in each study area as indicative of events with the potential for 5-minute concentrations of greatest interest in

[25] Data were stratified by two times of day (daytime - 6AM to 8PM; nighttime - all other hours) and four seasons (winter - December, January, February; spring - March, April, May; summer - June, July, August; and fall - September, October, November). Additionally, as the interest of this evaluation is occurrences of relatively higher concentrations during times of day and seasons when people are most likely to encounter them and is not regarding annual variability, the three years of data for each location are pooled before stratification.

this REA.[26] The model estimates are, for the most part, similar in magnitude to the monitor measurements (Figure 3-4). Across the study areas, the closest fit of the higher-concentration estimates to the monitor measurements for Fall River occurs in the spring, while concentrations for the summer and fall seasons appear to be somewhat over- and under-predicted by the model, respectively. In Indianapolis, the highest monitor concentration events in the spring are not reflected in the model estimates, while they may be somewhat over-predicted in the summer and fall seasons. In Tulsa, the higher concentration events observed at the monitor are not reflected by concentrations predicted by the model for any of the three seasons.

[26] The complete set of graphs for this evaluation considering all seasons and monitors are provided in Appendix K.

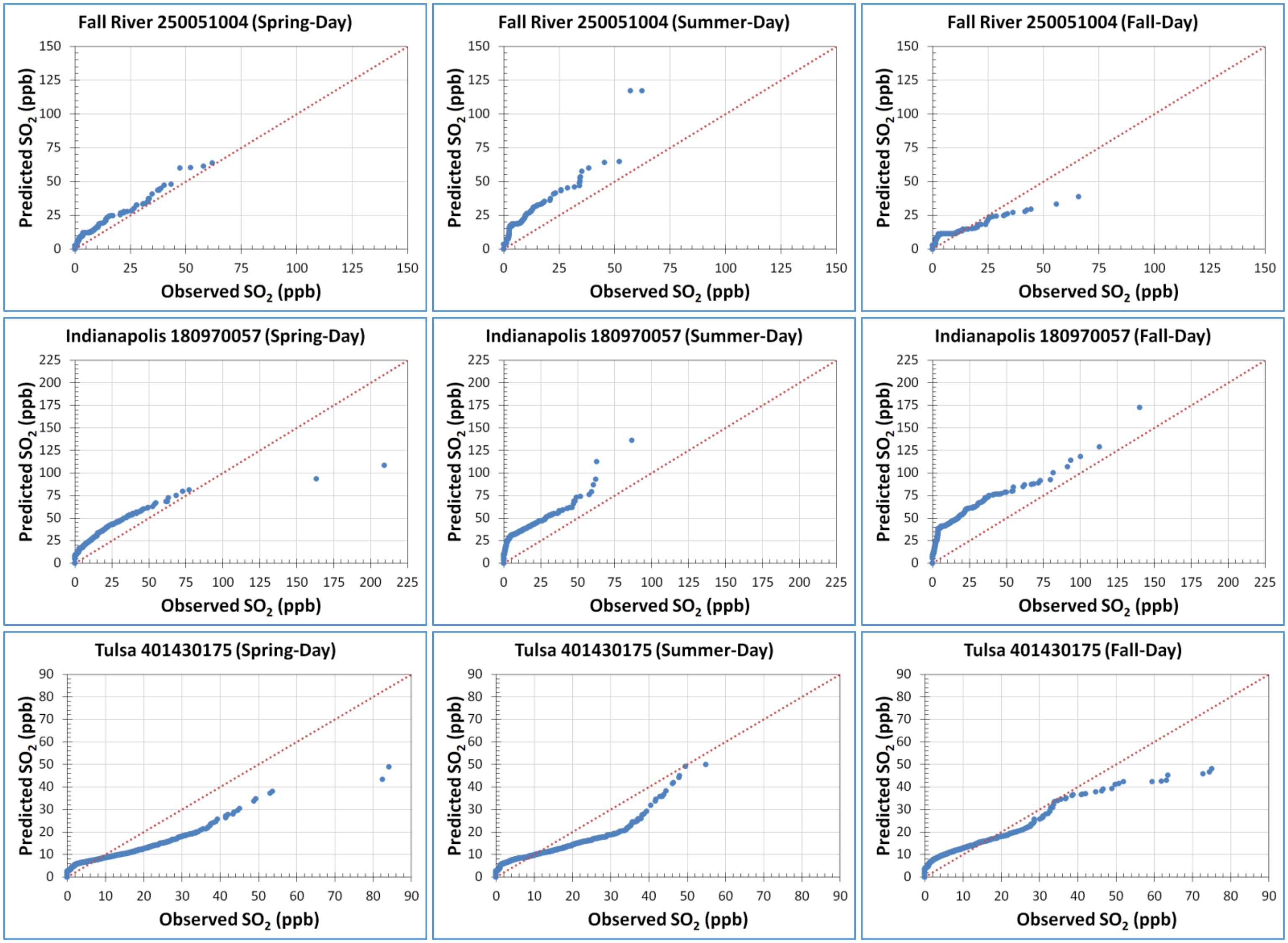


Figure 3-4. Comparison of AERMOD predicted SO_2 concentrations (y-axis) with observed air monitor SO_2 concentrations (x-axis) during daytime of the three warmer seasons at the highest design value monitor in each study area.

3.3 SELECTION OF AIR QUALITY RECEPTORS FOR EXPOSURE MODELING DOMAIN

As described above, the air quality modeling was done at a fine spatial scale that in some locations included receptor cells as small as 100 m by 100 m. Thus, the air quality modeling domains (Appendix C) included thousands of air quality receptor points, many more than considered practical for use by APEX in estimating exposures. APEX simulations were performed at a census block level, which, combined with the thousands of air quality receptors each considering the full 5-minute time-series of concentrations, presented computational challenges. In addition, the spatial range of the modeled air quality receptors extended outwards beyond areas expected to be influenced by the major sources present in each study area. Thus, the number of air quality receptors included in the exposure modeling was reduced to a more practicable number (i.e., fewer than 2,000) while still retaining the modeled receptors having the highest design value in the particular study area.

The approach used to define the exposure model domain within the air quality modeling domain in each study area, along with the number of air quality receptor sites included in the exposure modeling domain, is as follows:

- **Fall River:** Hourly SO_2 concentrations in ambient air were estimated at receptor sites defined by a 500 m grid. For the exposure modeling, we selected receptor sites that fell within 10 km of the Brayton EGU (latitude (lat) 41.709989, longitude (lon) -71.192441) and within 10 km of the continuous 5-minute monitor (lat 41.69, lon -71.17), which yielded 1,494 air quality receptors covering a land area of approximately 375 km^2.
- **Indianapolis:** Hourly SO_2 concentrations in ambient air were estimated at receptors defined by a receptor grid ranging from outside to inside at 2 km, then 1 km, 500 m, 250 m, and 100 m near the two major sources. For the exposure modeling, we selected receptor sites that fell within 10 km of the two major sources (Citizen Thermal: lat 39.762800, lon -86.166800; IP&L Harding: lat 39.7119, lon -86.1975) and all receptors within 10 km of Quemetco (lat 39.755391, lon -86.300155) and within 10 km of Indianapolis International Airport (lat 39.716809, lon -86.296127).[27] The finest scale grid concentrations retained were those falling within a 500 m interval, which yielded 1,917 air quality receptors covering a land area of approximately 675 km^2.
- **Tulsa:** Hourly SO_2 concentrations in ambient air were estimated at receptors defined by a receptor grid ranging from outside to inside at 1 km, 666.67 m, 500 m, 250 m, and 100 m near the two major sources (West Refinery: lat 36.139140 lon -96.025440; East Refinery: lat 36.11705271, lon -96.00477176). For the exposure modeling, we selected receptor

[27] Emissions from the Indianapolis International Airport were not explicitly modeled to remain consistent with the modeling performed for Indiana's SIP for the Marion County SO_2 nonattainment area; however, the exposure modeling domain was expanded using this source location to make this study area more representative of a large urban population.

sites that fell within 10 km of these two sources and receptor sites within 10 km of monitor 401431127 (lat 36.20, lon -95.98).[28] With the exception of 24 receptors modeled at a 100 m scale (retained in order to retain locations with the highest model-estimated DVs), the finest scale grid concentrations retained were those falling within a 500 m interval, which yielded 1,389 total air quality receptors covering a land area of approximately 550 km^2.

These exposure modeling domains for the three study areas are shown, with adjusted air quality per section 3.4 below, in Figures 3-6 through 3-8.

3.4 AIR QUALITY ADJUSTMENT TO CONDITIONS MEETING THE CURRENT STANDARD

The exposure and risk analyses were conducted for air quality adjusted to just meet the current primary SO_2 standard. Use of this adjusted air quality surface is most appropriate to quantitatively evaluate the associated exposures and health risks in this REA (section 2.2). As described in the REA Planning Document, a proportional approach was used to adjust ambient concentrations (not the modeled emissions) in the 2009 REA. An analysis of ambient concentration data at that time demonstrated that the proportional adjustment of ambient concentrations is an appropriate approach to use to generate air quality that just meets a particular standard (Rizzo 2009). We analyzed recent air quality data in the REA Planning Document to evaluate this assumption for several candidate areas for the purpose of justifying the selection of this approach for use in this REA (U.S. EPA, 2017b, Figure 4-6 and Appendix C). The results of the air quality comparisons shown in the REA Planning Document were similar to what was observed previously (Rizzo, 2009).

We further refined these air quality analyses here to include the monitoring data from the three REA study areas. We also extended the time period considered to encompass the most recent year in which ambient air monitor concentrations had a 99th percentile daily maximum 1-hour concentration at or just below the level of the current standard (i.e., 75 ppb, the air quality scenario adjustment goal) and the past year in each study area that had the highest daily maximum 1-hour concentrations (i.e., evaluate a maximum range in ambient air concentrations to reasonably support the use of potentially high adjustment factors, where needed). Thus, evaluated data included recent ambient air monitoring data from 2015 and measurements from as far back as 1980. We also focused the analysis on the monitor having the highest recent design

[28] In addition to SO_2 emission from the two refineries and the glass plant, the emissions from the PSO Northeastern Power Station were used to estimate ambient SO_2 concentrations in the Tulsa exposure study area (Table 3-6). However, the exposure study area was not expanded to include receptors near the PSO because it is located approximately 40 km northeast of Tulsa (see Figure 3-3).

values in each of the three study areas and paired two years having the two highest design values and with two recent years having design values at or just below 75 ppb.

Figure 3-5 presents the results of this ambient air monitoring data comparison. Concentrations are linearly related across the wide range of concentrations and, in a few instances, exhibit proportionality across the majority of the concentration distribution (i.e., in addition to exhibiting linearity, the regression intercept equals zero). Based on the paired 99^{th} percentile daily maximum 1-hour concentrations, the potential upper range of adjustment factors supported by these comparisons would range from 2.2 to 3.1. However, there are instances of non-proportionality as has been described previously (2009 REA; U.S. EPA 2017b), including limited deviation from linearity, particularly at the upper percentiles, and the presence of statistically significant linear regression intercepts. Thus, based on these analyses, we used a largely proportional adjustment approach in this REA with a variation from the 2009 REA approach, as described below to account for deviations from proportionality.

The process of adjusting air quality to just meet a standard of interest begins with consideration of the design values (DVs) calculated at the various locations in the study area. When using a proportional adjustment approach, the highest DV is used to derive a single factor (*F*) to adjust the monitored concentrations across the study area. In each study area, *F* is then used to adjust all SO_2 concentrations in a study area by this factor to simulate just meeting the current standard. In the case of the SO_2 standard, this adjustment of air quality is based on three years of concentrations, which is consistent with the form for the current standard.

A variation of this approach to air quality adjustment is used in this assessment. This new approach attempts to better consider relative source contributions to the ambient air concentrations that may or may not change given the particular air quality scenario. For instance, in the Fall River study area, the influence of the Brayton EGU (a source having >100 tpy SO_2 emissions in the area) was accounted for by air quality modeling as a point source and the resulting surface of modeled air concentrations was combined with the set of concentrations that account for emission sources not modeled in the study area. In considering how to derive a concentration surface reflecting the hypothetical scenario of air quality conditions just meeting the current standard, we concluded that adjusting just the concentrations resulting from the EGU emissions alone (rather than the aggregate concentrations from the EGU and the mix of concentrations from the sources not modeled) would create a scenario that better reflected how air concentrations would change in response to actions performed to meet air quality standards.

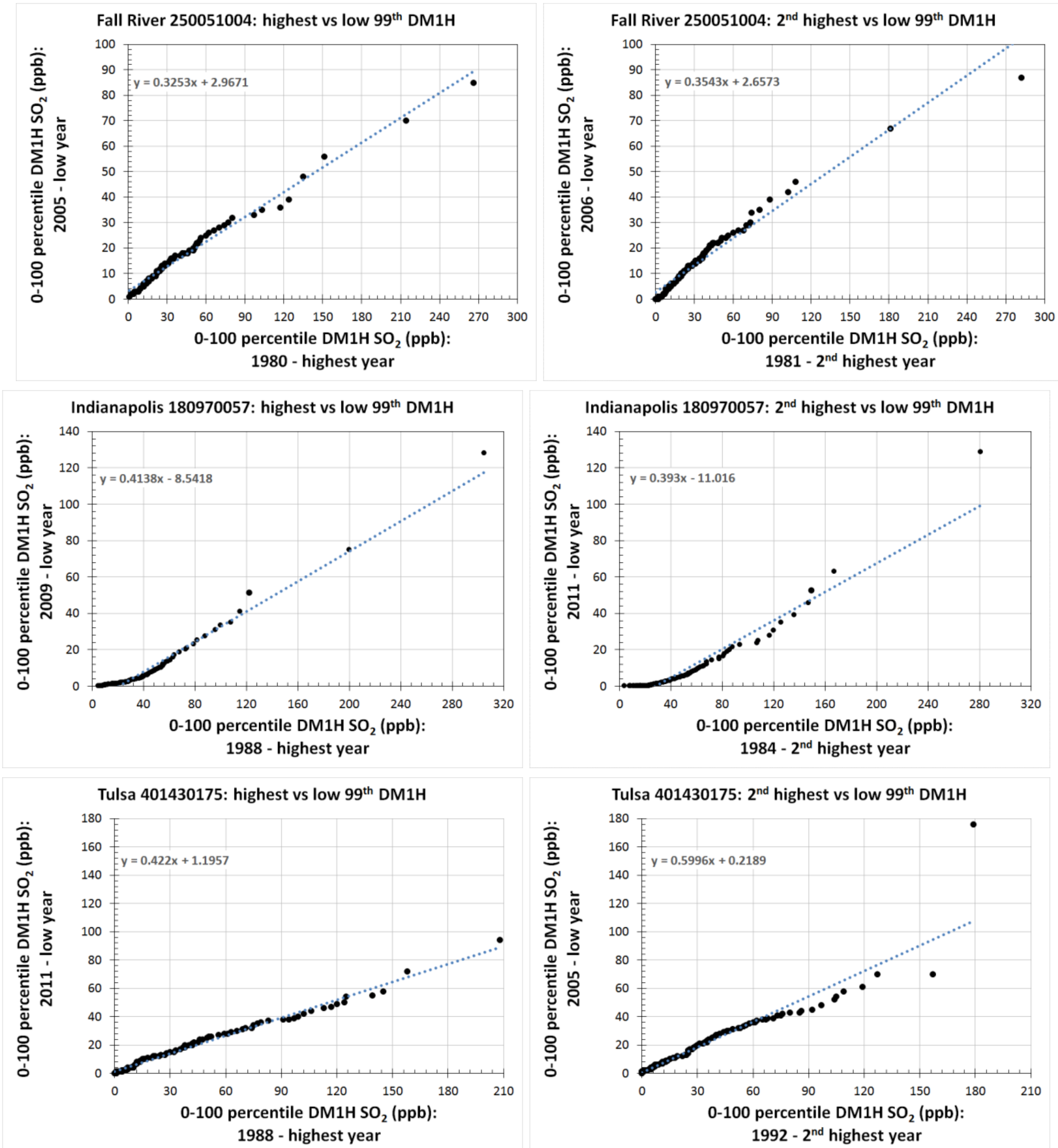


Figure 3-5. Comparison of ambient air measurements from high concentration years (x-axis) to low concentration years (y-axis) in the Fall River (top row), Indianapolis (middle row), and Tulsa (bottom row) study areas. Left column contains the year having the highest 99th percentile daily maximum concentration. Right column contains the year having the 2nd highest 99th percentile daily maximum concentration.

Accordingly, we applied this approach to the Fall River study area air quality, with the concentrations contributed from the EGU adjusted just enough such that the aggregate of these modeled concentrations and the concentrations not modeled just met the current standard at the air quality receptor having the highest design value. This concentration adjustment approach was also applied in a similar manner to the other two study areas, with a primary source (among the collection of sources modeled in these areas) identified for the air quality adjustment. Then, concentrations at air quality receptors that were contributed from all other sources were left unadjusted. In the Indianapolis study area, the air quality receptor concentrations contributed from each of the modeled sources were evaluated; the IP&L Harding Street Facility was identified as the primary contributor to most of the air quality receptors having the highest concentrations, particularly those within 10 km of the facility. A similar evaluation was done for the Tulsa study area; the West Refinery was identified as the primary contributor to the highest concentrations at air quality receptors in the study area.

The steps involved for this adjustment approach are summarized here. First, the maximum DV and associated air quality receptor (r_{max}) was identified among the DVs from the complete collection of modeled air quality receptors in each study area that comprise the exposure modeling domain. Then the following formula was used to calculate the single adjustment factor to be applied to the primary source concentrations (C_1), while considering the concentrations associated with the other sources (C_{oth}) as unchanged:

$$F = \frac{C_{1,rmax,2011} + C_{1,rmax,2012} + C_{1,rmax,2013}}{\{(75 \times 3) - (C_{oth,rmax,2011} + C_{oth,rmax,2012} + C_{oth,rmax,2013})\}} \qquad \text{Equation 3-1}$$

In order to have air quality just meet the current standard in each study area, the study area specific adjustment factor was used to adjust all hourly concentrations at each receptor as follows:

$$C_{std} = \frac{C_1}{(F)} + C_{oth} \qquad \text{Equation 3-2}$$

Table 3-8 contains the air quality receptor design values for each study area and the proportional adjustment factor that was applied to the concentrations that reflect the primary source emissions in each area in order to have concentrations just meet the current standard. Figures 3-6 to 3-8 show the air quality receptors in each study area and their respective design values following the above described approach for adjusting the hourly concentrations to just meet the current standard.

Table 3-8. Maximum SO_2 design values modeled at air quality receptors and associated proportional adjustment factors applied to primary source concentrations in each study area.

Study Area	Modeled Air Quality Receptor Maximum SO_2 DV (ppb)	Primary Source in Study Area	Proportional Adjustment Factor [a]
Fall River, MA	101.4	Brayton EGU	1.46
Indianapolis, IN	311.3	Harding EGU	4.21
Tulsa, OK	73.5	West Refinery	0.98

[a] The proportional adjustment factor is based on and applied only to the primary source contributing to the highest concentrations in the study area, while other source contributions as well as background concentrations are assumed to remain unchanged in approximating air quality conditions to just meet the current standard.

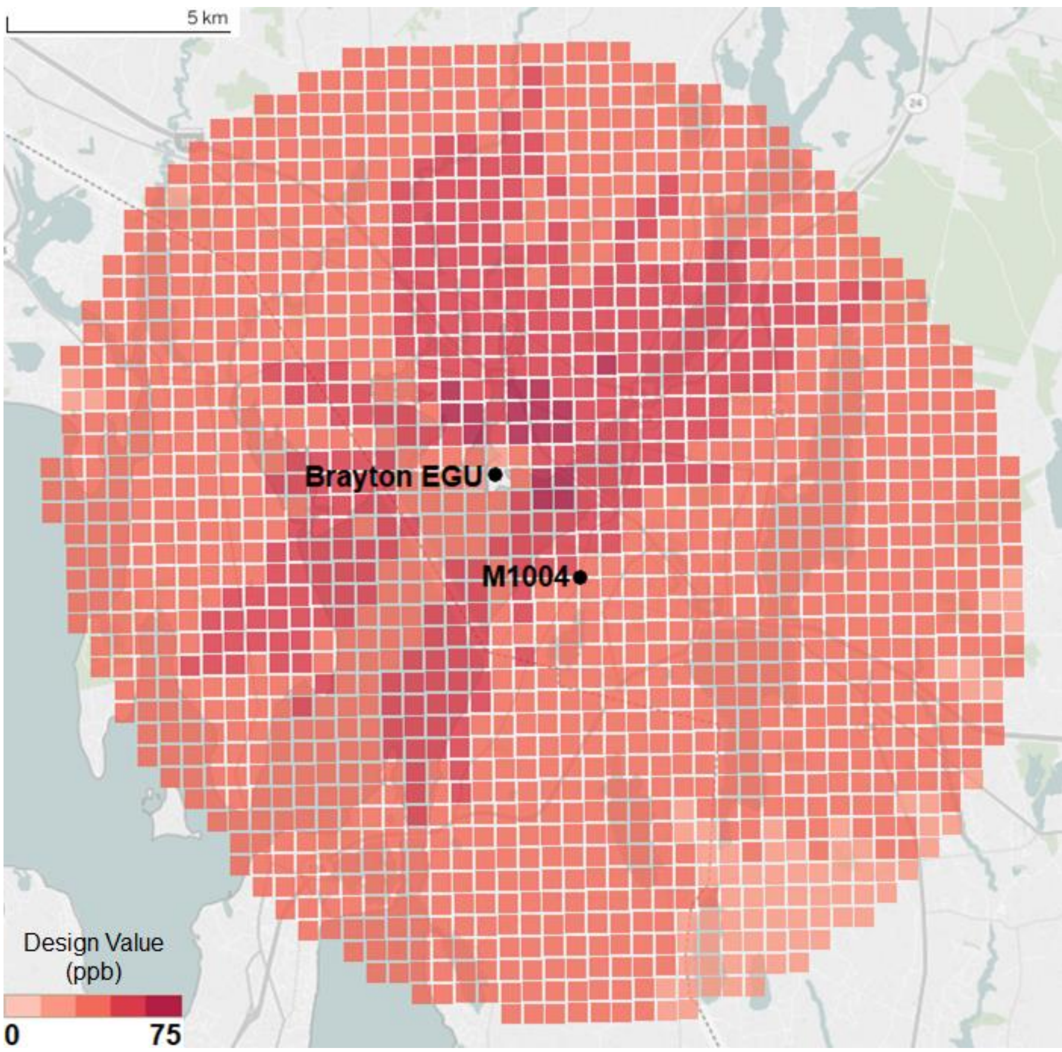


Figure 3-6. Location of air quality receptors, emission sources, and ambient monitors in the Fall River exposure modeling domain and receptor design values calculated from modeled hourly SO_2 concentrations adjusted to just meet the current standard.

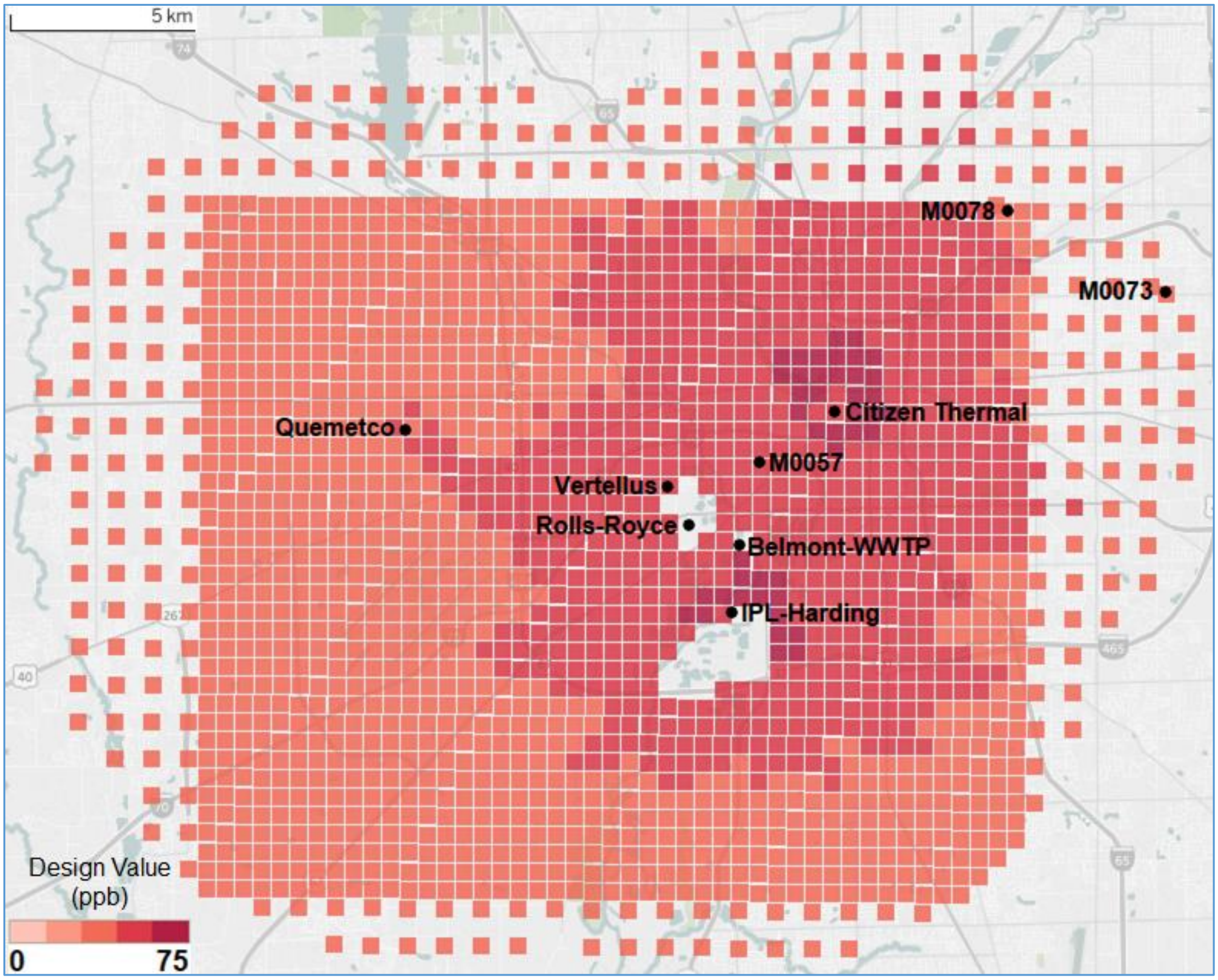


Figure 3-7. Location of air quality receptors, emission sources, and ambient monitors in the Indianapolis exposure modeling domain and receptor design values calculated from modeled hourly SO_2 concentrations adjusted to just meet the current standard.

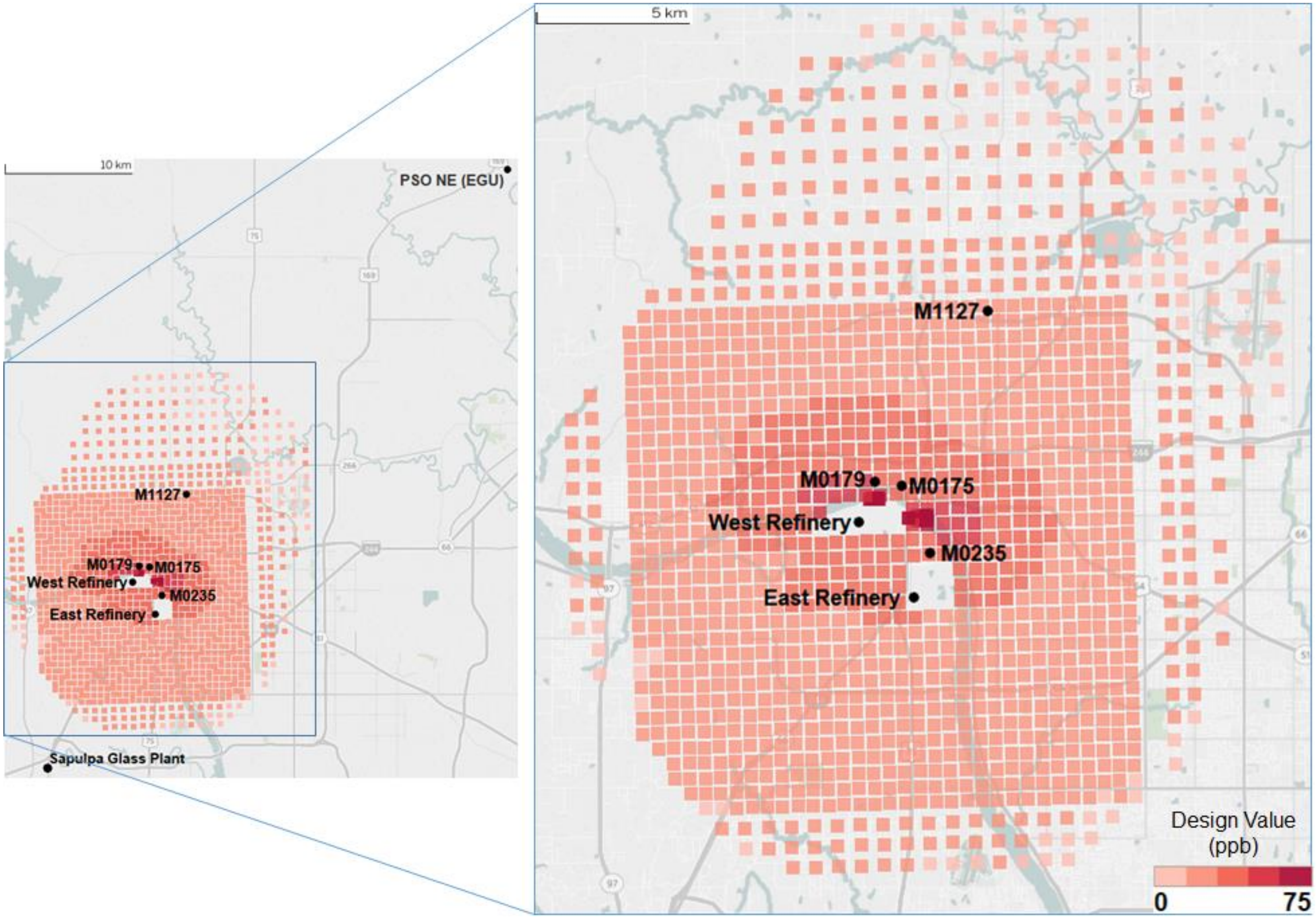


Figure 3-8. Location of air quality receptors, emission sources, and ambient monitors in the Tulsa exposure modeling domain and receptor design values calculated from modeled hourly SO_2 concentrations adjusted to just meet the current standard.

3.5 FIVE-MINUTE CONCENTRATIONS

In this assessment, we combined the fine-scale temporal characteristics of continuous 5-minute monitoring data local to each study area with the fine-scale spatial characteristics of hourly concentrations estimated by AERMOD. First, missing values within any monitoring data set were interpolated using the measured values immediately bounding the missing values. Then, where continuous 5-minute data were not available, an algorithm was constructed to randomly sample 5-minute concentrations from lognormal distributions that conform to the existing 1-hour average and maximum 5-minute measurements. Finally, the complete year pattern of 5-minute monitored concentrations was combined with the complete year pattern of hourly concentrations modeled at each receptor, based on matching the rank ordered 1-hour concentration distributions. The following section details how this was done, noting specifically where the approach differs from that described in the REA Planning Document.

3.5.1 Preparing Monitoring Data: Assessing Completeness & Filling Missing Values

Because there are years when the ambient air monitor did not report every hourly or 5-minute concentration and because APEX needs the complete time-series of 5-minute ambient air concentrations to estimate exposures, an approach was developed to approximate missing 5-minute values in the ambient air monitor data sets. As described in section 3.1 above, the study areas and years selected for this assessment corresponded to years in which the monitor datasets met completeness requirements for calculating a design value. This completeness requirement is typically applied to the hourly monitor concentrations and used for regulatory purposes. To best inform our estimation of 5-minute concentrations, we did not restrict the 5-minute concentrations using this completeness requirement for this assessment. Our intent in this REA was to utilize as much of the 5-minute measurements as was available in each study area.[29] From ambient air monitors in the three selected study areas, the following data sets containing 5-minute concentrations were available:

[29] For the hourly measurements, the following steps were taken: (1) a 75% completeness criterion was applied to each day monitored, with the monitored day considered valid if it contained measurements for at least 18 of the 24 hours; (2) the number of days within a quarter of the calendar year were evaluated, also using a 75% completeness criterion such that a monitored quarter was considered valid if there were at least 68-69 valid days and a year was considered complete if all four quarters were valid; (3) data were screened for outliers, such that hours in which a 5-minute max hourly value (AQS parameter 42406 and duration code 1) was reported and fell within a factor of 1 and 12 times the AQS hourly value (parameter 42401 and duration code 1) were kept. For the continuous 5-minute measurements, the screening for outliers was as follows: only 5-minute data with a corresponding hourly value in AQS (parameter 42401 and duration code 1) were kept and only 5-minute values with an hourly mean value less than 120% of the hourly value in AQS (parameter 42401 and duration code 1) were kept.

- **Fall River:** continuous 5-minute data were available for 2011 and 2012. For 2013, the maximum 5-minute concentrations within the hour were available.
- **Indianapolis:** continuous 5-minute data were available for 2011-2013.
- **Tulsa:** continuous 5-minute data were available for 2011-2013.

A simple approach was selected to estimate any missing 1-hour, maximum 5-minute, and continuous 5-minute concentrations within the ambient air monitor data sets listed in Table 3-9. We used PROC EXPAND (SAS, 2017) to interpolate between missing values, using the measured values that bound the missing data to estimate missing concentrations via the JOIN method (SAS, 2017). This approach fits a continuous curve to the data by connecting successive straight-line segments. While this approach does not directly calculate an average of the concentrations surrounding data gaps and generate a single concentration to use for all hours within a particular gap, the degree of variability assigned to concentrations within multi-hour gaps is limited. While more complex methods exist (e.g., autoregressive models) to perhaps increase the representation of variability that might be occurring within multi-hour data gaps, the performance of these simple methods is similar to complex methods when filling data sets having few (< 5-10%) missing values (Junger and de Leon, 2015).

To support the use of this method to substitute for missing values, we evaluated monitoring data available in the three study areas. Table 3-9 provides the number of missing values within each 1-hour, maximum 5-minute, or continuous 5-minute across the 3-year period and the percentage that number is of the number of values in a full dataset. There were very few instances where the gap of missing data spanned several hours to days. The percentage of the total dataset values that were missing was at or less than 5% in nearly all instances when considering the Fall River and Tulsa Study areas. In contrast, the percentage of missing data for some of the ambient air monitors in the Indianapolis study area was greater. For example, Indianapolis monitor 18090073 had 40-60% of hours missing concentrations in both the continuous 5-minute data set and the maximum 5-minute and 1-hour data sets, and thus was not considered useful in subsequent assessment calculations (and was not used in further analyses). Indianapolis monitor 18090057 also had a large percentage of missing continuous 5-minute data for two of the years (25-58% missing), although it still had robust reporting of the maximum 5-minute and 1-hour data (1-3% missing) for each of the three monitor years. Concentration reporting from the Indianapolis monitor 18090078 was fairly complete considering either data set and for all three years (4-9% missing). We recognize that Indianapolis monitor 180970057 exceeds the above recommendation of having somewhere between 5 to 10% missing data when using the simple interpolation to estimate 5-minute concentration, however, we decided that use of continuous 5-minute concentrations from the local monitor was better than use of a surrogate monitor from Detroit to represent variability in 5-minute concentrations, as was done in the draft REA.

For each of the three study areas, we used the PROC EXPAND interpolation approach to fill the missing continuous 5-minute concentrations, with the exception of the Fall River monitor 250051004 in 2013 that only reported 5-minute maximum and 1-hour concentrations. Therefore, a second approach was employed to estimate within hour 5-minute concentration variability for this Fall River monitor (section 3.5.2). A complete set of 1-hour and continuous 5-minute data was first needed to apply this second approach. To estimate missing 1-hour and continuous 5-minute data for the 2013 Fall River monitor, PROC EXPAND used their respective measured concentrations to interpolate the missing values. Because of the dependence of 1-hour concentrations and maximum 5-minute concentrations,[30] the following steps were used for estimating missing maximum 5-minute concentrations:

- Using PROC EXPAND, estimate the missing 1-hour concentrations for each monitor and year;
- Calculate peak-to-mean ratios (PMRs) using the measured 1-hour and maximum 5-minute concentrations;
- Using PROC EXPAND, estimate the missing PMR values for each monitor and year;
- Calculate missing maximum 5-minute concentrations by multiplying the complete set of PMRs by their corresponding 1-hour concentrations.

[30] PROC EXPAND could have been used to estimate the missing maximum 5-minute concentrations based on using the measured values; however, this was not done because these simulated 5-minute values would not have been entirely consistent with the estimation of missing hourly concentrations. This lack of consistency would lead to PMRs that fall outside of the mathematically acceptable range (i.e., $1 \leq PMR \leq 12$). For this reason, measurement related PMRs were used for the interpolation of missing PMR (with a restriction to remain between 1 and 12) to ultimately estimate reasonable maximum 5-minute concentrations. The minimum ratio is 1 because the highest 5-minute concentration in an hour could never be less than the hourly mean. The maximum ratio is 12 because if the maximum 5-minute concentration (max5) was the only measured non-zero value (i.e., all other 11 5-minute measurements are 0), the hourly mean would be $(max5 + (11 \times 0))/12$ or simply $max5/12$, thus effectively yielding a $PMR = max5/(max5/12) = 12$.

Table 3-9. Percent of missing values in the hourly and 5-minute ambient air monitoring data sets for the three study areas (2011-2013).

Study Area	Monitor ID	Year	Continuous 5-minute data		1-hour and 5-minute maximum data	
			% Missing	Days/Year < 75% complete	% Missing	Days/Year < 75% complete
Fall River	250051004	2011	3.5	4	-	-
		2012	2.9	2	-	-
		2013	-	-	4.7	7
Indianapolis	18090057	2011	58.0	33	1.2	2
		2012	25.5	10	2.1	5
		2013	11.6	25	2.8	9
	18090078	2011	9.2	32	8.0	31
		2012	4.5	10	4.3	9
		2013	8.2	26	7.4	22
	18090073	2011	42.6	208	41.4	202
		2012	52.6	281	51.4	272
		2013	64.6	311	63.8	304
Tulsa	401430175	2011	1.2	2	-	-
		2012	1.1	3	-	-
		2013	2.6	9	-	-
	401430179	2011	-	-	-	-
		2012	-	-	-	-
		2013	3.2	12	-	-
	401430235	2011	2.7	10	-	-
		2012	3.3	12	-	-
		2013	1.6	4	-	-
	401431127	2011	1.3	5	-	-
		2012	7.3	31	-	-
		2013	2.3	7	-	-

The symbol "-" indicates there were no data needed for this evaluation because there were adequate continuous 5-minute data available or there were no data available.

3.5.2 Estimating Continuous 5-minute Concentrations at Monitor Having Only 1-hour Average and Hourly Maximum 5-minute Data

In this assessment, we are interested in estimating 5-minute exposures using the complete time-series of 5-minute ambient air concentrations for each year. We are also interested in utilizing, to the maximum extent possible, the local ambient air measurements to inform this estimation. As described above, there were no 5-minute continuous measurements available for one year (2013) for the Fall River study area. Based on the ambient air monitoring data that were available (i.e., 1-hour average and maximum 5-minute concentrations within each hour) and knowing that air pollutant concentrations are typically lognormally distributed (Kahn, 1973), an approach was developed to estimate the eleven other 5-minute concentrations occurring within each hour in the year for which continuous 5-minute measurements were not available. While early studies (e.g., Larsen, 1977) have developed models to estimate a few of the upper percentiles of a concentration distribution using relationships between peak concentrations and time-averaging (e.g., estimate a 2nd highest 1-hour from the 2nd highest 8-hour), they are not considered directly applicable to estimating a complete time-series of continuous 5-minute concentrations in a year (i.e., 105,120 values). We also note that there are maximum 5-minute monitored concentrations associated with instances where the hourly concentrations are reported, which already provides appropriate values for important peak 5-minute concentrations. Because the Fall River study area had continuous 5-minute data available for two of the years of interest, while also needing an approach to estimate continuous 5-minute concentrations for 2013, the 2011-2012 Fall River continuous 5-minute data served as a case study for developing and evaluating this approach.

We first evaluated the 5-minute data set to confirm lognormal distributions would be appropriate to fit the twelve measured 5-minute values in each hour and to determine the parameters associated with that distribution. Using the set of continuous 5-minute monitor data in Fall River (2011-2012), where all twelve[31] 5-minute measurements within an hour were available, data were categorized by their 1-hour average concentrations and their peak to mean ratios (i.e., PMRs, the maximum 5-minute concentration divided by the 1-hour average). This categorization was done because the 2009 REA analyses indicated a relationship between the magnitude of hourly SO_2 concentrations and the magnitude of the PMRs, consistent with conclusions made regarding this relationship (Singer, 1961). For the hourly concentrations, bins of 10 ppb increments were used to categorize hourly concentrations upwards from 0 through 80 ppb, with a final bin containing all concentrations above 80 ppb (yielding a total of 9 hourly concentration bins). PMR was categorized by 0.5 increments from 1 to 2, then in whole units from 2 to 4, ending with a final PMR bin of ≥ 4 (yielding a total of 5 PMR bins).

Then, we used PROC CAPABILITY (SAS, 2017) to evaluate the fit of eight statistical distribution forms[32] for both the varying hourly concentration and PMR binned continuous 5-minute data. Distribution fits were evaluated using four goodness-of-fit statistics: Kolmogorov Smirnov, Cramer von Mises, Anderson Darling, and Chi-Square (SAS, 2017). Best fit distributions were selected based on having the lowest p-value (or highest critical value) in the collection of fit statistics. For the low 1-hour concentration binned data (e.g., 0 to <10 ppb, 10 to <20ppb), normal distributions were found to have the best statistical fit, while for higher 1-hour concentration binned data, lognormal distributions had the best statistical fit (along with a few

[31] One hour has 12 five-minute periods (60/5=12), thus there are a total of twelve 5-minute concentrations possible within an hour.

[32] Distributions evaluated were normal, lognormal, Weibull, gamma, Pareto, exponential, beta, and Rayleigh.

having gamma and Weibull distributions as the most reasonable fit). This was not entirely unexpected given that some of the distribution types could not be fit to the binned data set (e.g., the number of samples in some of the bins was too small, the prevalence of concentration values of 0). Overall, the results indicate the within-hour 5-minute concentrations are generally consistent with a lognormal distribution, particularly considering high concentrations of interest, and that a lognormal distribution can be used to reasonably approximate the missing eleven within-hour 5-minute concentrations.

To do so, the parameters of all the fitted normal distributions were transformed to lognormal terms (geometric means and standard deviations) (Casella and Berger, 2002) and combined with the suite of parameters estimated for all of the fitted lognormal distributions. Series of twelve 5-minute concentrations were randomly sampled from these distributions for thousands of iterations, creating a new data set consisting of a distribution of thousands of datasets of twelve 5-minute concentrations, each lognormally distributed and having their own hourly average concentration and PMR. Individual sets of twelve 5-minute concentrations were then divided by their respective 1-hour average concentrations to create sets of normalized 5-minute concentrations (estimated concentrations), and then categorized by their PMR in 0.1 increments. For method validation, a test data set was created from the 2011-2012 Fall River monitor data, using only the observed 1-hour average and maximum 5-minute concentrations. From the data set of estimated concentrations, a set of twelve mean normalized[33] 5-minute concentrations were then randomly assigned to each 1-hour/maximum 5-minute concentration in the test data set and were linked using the same categorization of PMR in 0.1 increments. Finally, the within-hour continuous 5-minute concentrations were calculated for each hour by multiplying the observed 1-hour average by the normalized twelve 5-minute concentrations.[34]

The complete set of estimated 1-hour mean, 5-minute maximum, and continuous 5-minute concentrations were compared with the respective metric in the monitoring dataset. Figure 3-9 illustrates the relationship, indicating excellent reproducibility of the original 1-hour (top panels) and maximum 5-minute concentrations (middle panels) and reasonable agreement between the estimated and measured 5-minute continuous concentrations (bottom panels). Table 3-10 provides summary statistics for comparison to further support the relationship.

[33] All twelve 5-minute concentrations occurring within an hour were divided by that hourly 1-hour average concentration.

[34] Where needed, a small downward or upward adjustment was applied to the suite of 5-minute concentrations to ensure the modeled values had a 1-hour average and maximum 5-minute concentration consistent with the monitoring measurements. The approach was designed to precisely replicate the 1-hour average and its associated variability of all 12 within hour 5-minute concentrations, thus there are a few instances where the estimated and measured 5-minute maximum deviated slightly from one another (Figure 3-9, middle panel).

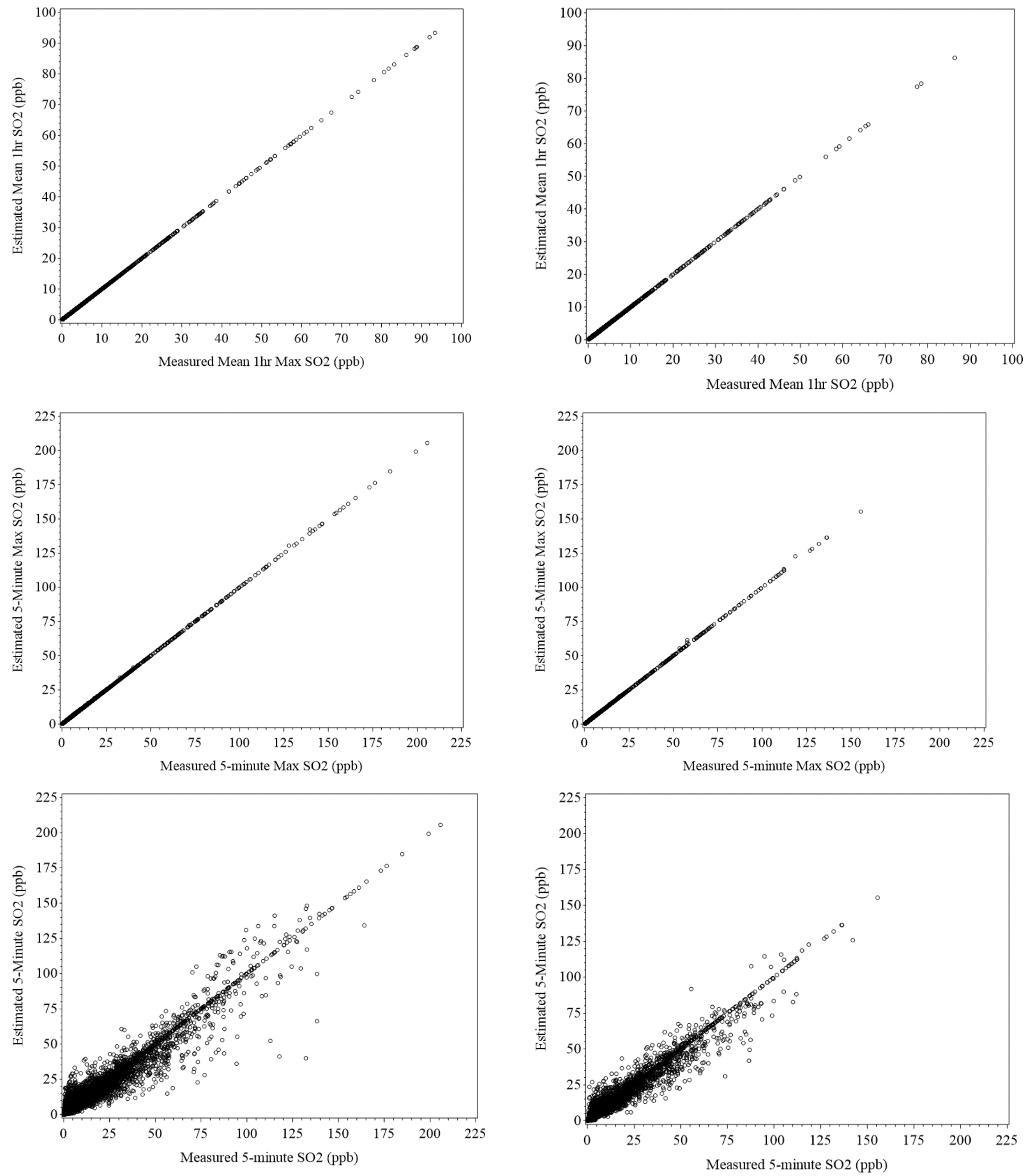


Figure 3-9. Comparison of estimated to measured SO_2 concentrations in ambient air in Fall River monitor 250051004: 1-hour average (top panels), maximum 5-minute (middle panels) and continuous 5-minute (bottom panels) for 2011 (left panels) and 2012 (right panels).

Table 3-10. Descriptive statistics and correlations associated with measured and estimated 1-hour average, maximum 5-minute, and continuous 5-minute SO_2 concentrations, Fall River (monitor 250051004), 2011-2012.

Variable	Year	Data Set	SO_2 Concentrations (ppb) N	Mean	Std Dev	Minimum	Maximum	Correlation (r)
1-hour average	2011	Estimated	7728	3.01	5.97	0.09	93.4	1.00000
		Measured	7728	3.01	5.97	0.09	93.4	
	2012	Estimated	8404	2.43	4.27	0.11	86.3	1.00000
		Measured	8404	2.43	4.27	0.11	86.3	
Maximum 5- minute	2011	Estimated	7728	5.62	14.11	0.2	205.7	0.99999
		Measured	7728	5.59	14.11	0.2	205.7	
	2012	Estimated	8404	4.04	9.71	0.2	155.5	0.99996
		Measured	8404	4.01	9.70	0.2	155.5	
Continuous 5-minute	2011	Estimated	92736	3.01	7.03	0.02	205.7	0.97516
		Measured	92736	3.01	7.24	0	205.7	
	2012	Estimated	100848	2.43	4.94	0.03	155.5	0.97922
		Measured	100848	2.43	5.11	0	155.5	

3.5.3 Estimating 5-minute Concentrations Across Study Areas

In the following sections we discuss the approach used to estimate 5-minute concentrations across the exposure modeling domain encompassing each study area (section 3.5.3.1), summarize the estimated 5-minute concentrations in relation to available ambient air 5-minute measurements (section 3.5.3.2), and include a comparison of estimated concentrations with ambient air monitor measurements considering the occurrence of concentrations at or above concentrations of interest during times of greater exposure potential (section 3.5.3.3).

3.5.3.1 Combining 5-minute Monitor Measurements with 1-hour AERMOD Receptor Estimates

The complete temporal profile of each of the three years of continuous 5-minute monitor data developed using the above approach(es) was used to approximate the within-hour variation in 5-minute concentrations at each AERMOD air quality receptor site in each study area. The approach used in this REA to combine the monitor data with the modeled hourly estimates is a slight variation of that described in the REA Planning Document.[35] We have adjusted the proposed approach in the REA Planning Document to better reflect instances where the ambient

[35] For the REA Planning Document, we originally proposed to match by consecutive hour, i.e., using the complete calendar years of hourly concentrations for both the ambient monitor and each air quality receptor. Then, each within-hour distribution of twelve 5-minute concentrations from the monitor would be adjusted using a multiplicative factor derived from the ratio of the 1-hour average concentrations (i.e., modeled divided by measured) (see REA Planning Document, Equation 4-4).

air monitor may capture a high concentration event that may not necessarily occur at the same clock time at a modeled air quality receptor that is located at a distance from the monitor. Events such as these would result from varying lateral or vertical transport of pollutant plumes that may not necessarily be captured by the air quality modeling,[36] affecting both the temporal and spatial characteristics of the air quality surface.

Considering this, the calendar-based approach (described in the REA Planning Document) could result in a mismatching of times when peak concentration occurs across the spatial domain and thus lead to potentially erroneous distributions of 5-minute concentrations. For this REA, we linked the high concentration events occurring in both the monitor data set and the modeled hourly estimates at air quality receptors by ranking their respective 1-hour concentration distributions. Thus, all low 1-hour concentrations at each modeled air quality receptor will be linked to the distribution of 5-minute concentrations that occur during low 1-hour concentrations measured at the monitor, and, in a similar fashion, all high hourly concentration events will be appropriately linked, irrespective of clock hour. A similar equation to that provided in the REA Planning Document that replicates the pattern of the monitored 5-minute values in an hour by scaling the 5-minute values so their hourly averages are equal to the AERMOD predictions for that hour (Equation 3-3) is described here:

$$Y_{s,r,i} = \frac{Y_{s,r}}{\frac{1}{12}\sum_{i=1}^{12} X_{r,i}} X_{r,i} \qquad \text{Equation 3-3}$$

where,

$X_{r,i}$ = the ith 5-minute value (ppb) at the monitor, having 1-hour ranked concentration *r*

$Y_{s,r}$ = the 1-hour AERMOD value (ppb) at location s, having 1-hr ranked concentration *r*

$Y_{s,r,i}$ = the ith 5-minute value (ppb) having 1-hr ranked concentration *r*, at location *s*

s = AERMOD prediction point in space

r = rank ordered 1-hour concentration, r = 1, 2, …, 8760 (or 8784 for leap years)

i = sequence of 5-minute values within the hour, i = 1, 2,..., 12.

Thus, the complete year distribution of continuous 5-minute concentrations was applied to the modeled receptors using the complete time-series of hourly scaling factors (unique to each receptor) to yield the time-series of 5-minute SO_2 concentrations (e.g., n= 12×24×365 = 105,120

[36] There is variation in the emissions and meteorological data input to the model relative to the actual emissions and meteorology. For example, it is possible that, given the limited number of meteorological stations and their geographic locations relative to the hundreds of receptors modeled across a 200 km^2 study area, the actual local fine scale weather patterns will not all coincide in time and space.

values) at every air quality receptor in each study area. Effectively, all spatial gradients that may exist for each hour across the study area are maintained; the 5-minute monitoring data only add a finer scale to the within-hour temporal variability. Because the ranked concentration distributions for each modeled air quality receptor may have a differing order of actual clock hours, it is likely that the within-hour 5-minute concentration variability (and hence maximum 5-minute concentrations) differs across the air quality receptors when considering the same clock hour. This is considered a reasonable and realistic outcome of using this approach.

For instances where a study area has more than one ambient air monitor (i.e., Indianapolis and Tulsa), modeled receptors were linked with 5-minute concentration data from the nearest monitor. Again, all spatial gradients that may exist within each hour across the study area are maintained and it is likely that there is differing within-hour 5-minute concentration variability and occurrence of maximum 5-minute concentrations across the air quality receptors when considering the same clock hour. The assignment of monitor to modeled air quality receptors is as follows:

- **Fall River:** all air quality receptors were linked to 5-minute concentrations from the single ambient air monitor in the study area (250051004).
- **Indianapolis:** monitor 180970057 is located between the two largest sources (Harding and Citizens Thermal) and is considered to best represent local source related 5-minute concentration variability. The 5-minute concentrations from this monitor were linked to air quality receptors within 10 km of Harding and 5 km within Citizens Thermal, i.e., those receptors potentially having a strong local source influence. All other receptors used monitor 180970078 to represent air quality receptors not having a strong local source influence on 5-minute concentrations. Monitor 180970073 is considered outside of the exposure modeling domain and had a large percent of missing data, thus these data were not used at this time.
- **Tulsa:** monitor 401430175 is closest to the West Refinery and monitor 401430235 is closest to the East Refinery. These monitors are considered to best represent local source related 5-minute concentration variability. Based on the spatial pattern of DVs, concentrations from monitor 401430175 were linked to air quality receptors within 10 km of the West Refinery and concentrations from monitor 401430235 were linked to receptors within 5 km of the East Refinery. All other receptors used monitor 401431127 to represent air quality receptors not having a strong local source influence on 5-minute concentrations. Monitor 401430179 is proximal to monitor 401430175, although further from the West Refinery. This monitor only has data for 2013 and was not used to estimate 5-minute concentrations at this time.

3.5.3.2 Summary of Estimated 5-minute Concentrations Across Study Areas

After estimating the continuous 5-minute concentrations at each air quality receptor location, the distributions of these 5-minute concentrations were compared to those of the 5-minute ambient air measurements in each study area. To do so for this comparison, the ambient air monitor concentrations in each study area were first adjusted proportionally using the single factor derived from the maximum monitor design value to reflect conditions that would just meet the current standard. As such, the adjusted ambient air concentrations from the monitor having the highest design value would hypothetically represent a distribution of the highest concentrations in a study area among the monitored data set.[37]

We summarized the monitor continuous 5-minute concentrations by identifying the 90th and 99th percentiles of the distribution and selecting the maximum 5-minute concentration. The estimated continuous 5-minute concentrations at the air quality receptor sites were also summarized by considering the upper percentiles of the distribution. The 90th and 99th percentiles of the distribution, along with the maximum 5-minute concentration, were identified at each modeled receptor location. Because there were over a thousand air quality receptors within each study area, we consolidated each of these statistics to a new set of statistics. We still focused on the 90th and 99th percentiles of the distribution and the maximum 5-minute concentration, however now we considered the distribution of each of these upper percentile concentrations across the entire set of air quality receptors. For example when considering the *maximum* 5-minute concentrations, the maximum of all the *maximum* 5-minute concentrations (i.e., the single highest air quality receptor concentration considering the entire study area), the 99th percentile of all *maximum* 5-minute concentrations (i.e., 1% of the complete set of modeled receptors have a *maximum* 5-minute concentration greater than this value), and the 90th percentile of all *maximum* 5-minute concentrations (10% of the complete set of modeled receptors have a *maximum* 5-minute concentration greater than this value) would be presented. This summary sequence would then follow for the other two statistics (the upper percentile distribution of all *90th* and *99th* percentile 5-minute concentrations from the collection of receptors) generated from the collection of air quality receptors, which are provided in Tables 3-11 through 3-13.

There is reasonable agreement at the upper percentiles between the adjusted monitored concentrations and the estimates developed for the receptor sites, particularly considering the 99th percentile and maximum values in the Fall River and Tulsa study areas (Table 3-11 and 3-13). For example, the range in particular percentile concentrations (e.g., the 90th, 99th, and maximum of the estimated maximum percentile 5-minute concentrations across all receptors) estimated for the model receptor locations bound the measured 5-minute concentrations quite well (e.g., maximum 5-minute concentrations for 2011 and 2012 in the Fall River study area). In some instances, the range of upper percentile concentrations for the model receptor sites extends above the monitor upper percentile concentrations (e.g., the 99th percentile concentrations in Fall

[37] Therefore, the maximum hourly design value for both the ambient monitor and modeled receptor would be 75 ppb, making the two sets of data more compatible.

River for 2012 and 2013). In other cases, the range of the receptor upper percentile concentrations is below the monitor upper percentile concentrations (e.g., the maximum 5-minute concentrations in Fall River for 2013).

For most percentiles of the concentration distribution for the Indianapolis study area, the receptor concentrations are greater than the monitor-based concentrations. This may be related to the approach for estimating concentrations associated with source emissions not explicitly modeled (Table 3-7).[38] Given that the range of maximum 5-minute concentrations estimated at the receptor locations (e.g., 452-642 ppb for the first year) extends above that of the monitor (355 ppb), it is possible that, even when adjusted for just meeting the current standard, these maximum 5-minute concentrations at modeled receptor sites appear somewhat high. However, we also note that there are situations where the estimated maximum 5-minute concentrations at receptor sites were well below that of the monitor (e.g., receptor concentrations peaked between 167-205 ppb compared to monitor concentration of 369 ppb for year two of the simulation). It may also be that the numerous receptors situated in close proximity to the largest emissions sources in the area are representing hourly (and hence 5-minute) variability not reflected by the monitors. In the absence of having monitors at all the receptor sites to confirm this, the upper range of predicted concentrations across each of the study areas remain as an important uncertainty.

[38] In this same analysis performed for the draft REA (which relied on a different approach for estimating concentrations associated with source emissions not explicitly modeled), concentrations estimated at the lesser primary source influenced receptors were less than that observed for monitor 180970078. While the draft REA used a surrogate monitor to approximate 5-minute variability in Indianapolis and this REA used the continuous 5-minute data from study area monitors, this change in approach does not appear to be a significant contributor to the differences between the concentration distributions (data evaluation not shown).

Table 3-11. Descriptive statistics for concentrations at monitors and concentrations estimated at air quality receptor locations, Fall River study area 2011-2013.

Unadjusted or Adjusted Values	Type of Statistic	2011	2012	2013 [a]
Monitor (250051004) 5-minute SO_2 Concentrations (ppb)				
unadjusted	p90	4	3	4
	p99	31	21	12
	max	206	156	206
adjusted [b]	p90	5	4	4
	p99	37	25	14
	max	241	182	241
Estimated 5-minute SO_2 Concentrations (ppb) at Air Quality Receptors				
adjusted [c]	p90p90	11	10	11
	p99p90	11	10	11
	maxp90	11	10	11
	p90p99	32	27	22
	p99p99	41	31	24
	maxp99	48	35	26
	p90max	183	129	121
	p99max	247	187	150
	maxmax	268	214	180

[a] For 2013, only the maximum 5-minute measurement concentrations were available in Fall River, even though this evaluation includes estimated continuous 5-minute concentrations for monitor 250051004.
[b] Adjusted concentrations were based on a monitor-based design value (adjustment factor =64/75 = 0.85).
[c] Adjusted concentrations were based on highest modeled air quality receptor and the primary source contribution to concentrations at that receptor (see section 3.4).
Abbreviations: pN= Nth percentile of 5-minute concentrations at monitor; pNpN = Nth percentile of the distribution of all study area receptor Nth percentile 5-minute concentrations. For example, p90 = 90th percentile of 5-minute concentrations at monitor and p90p99 = 90th percentile of the distribution of all study area receptor 99th percentile 5-minute concentrations.

Table 3-12. Descriptive statistics for concentrations at monitors and concentrations estimated at model receptor locations, Indianapolis study area 2011-2013.

Unadjusted or Adjusted Values	Type of Statistic	2011	2012	2013	2011	2012	2013
		Monitor 5-minute SO_2 Concentrations (ppb) [a]					
		Local Primary Source Influence (monitor 180970057)			Less Primary Source Influence (monitor 180970078)		
unadjusted	p90	4	5	5	5	6	5
	p99	22	38	34	29	31	37
	max	370	384	255	99	108	107
adjusted [b]	p90	3	4	5	5	6	5
	p99	21	37	33	29	30	36
	max	355	369	245	99	104	103
		Estimated 5-minute SO_2 Concentrations (ppb) at Model Receptors					
		Local Primary Source Influence			Less Primary Source Influence		
adjusted [c]	p90p90	21	19	20	18	18	18
	p99p90	23	21	22	20	19	19
	maxp90	25	23	23	20	19	20
	p90p99	52	54	53	44	45	44
	p99p99	61	62	61	45	46	45
	maxp99	68	66	67	45	47	45
	p90max	452	167	239	132	157	145
	p99max	534	188	286	132	157	145
	maxmax	642	205	343	135	157	145

[a] For all years monitored, continuous 5-minute measurement concentrations were available.
[b] Adjusted concentrations were based on a monitor-based design value (adjustment factor =78/75 = 1.04).
[c] Adjusted concentrations were based on highest modeled air quality receptor and the primary source contribution to concentrations at that receptor (see section 3.4).
Abbreviations: p90 = 90th percentile of 5-minute concentrations at monitor. p90p90 = 90th percentile of the distribution of all study area receptor 90th percentile 5-minute concentrations.

Table 3-13. Descriptive statistics for concentrations at monitors and concentrations estimated at model receptor locations, Tulsa study area 2011-2013.

Adjusted or Unadjusted Values	statistic	2011	2012	2013	2011	2012	2013	2011	2012	2013
	Monitor 5-minute SO_2 Concentrations (ppb) [a]									
		Local Primary Source Influence (401430175)			Local Primary Source Influence (401430235)			Less Primary Source Influence (401431127)		
unadjusted	p90	15	11	7	2	1	1	2	2	1
	p99	50	42	33	17	5	7	8	5	4
	max	154	152	123	114	77	50	67	33	84
adjusted [b]	p90	20	15	10	3	1	1	2	2	2
	p99	68	57	45	23	7	10	11	7	5
	max	210	207	168	155	105	68	92	46	114
Estimated 5-minute SO_2 Concentrations (ppb) at Model Receptor Locations										
		Local Primary Source Influence			Local Primary Source Influence			Local Primary Source Influence		
adjusted [c]	p90p90	10	10	8	7	7	6	5	5	5
	p99p90	29	24	14	12	10	7	6	6	5
	maxp90	41	37	17	13	11	8	6	6	5
	p90p99	41	34	23	35	28	22	16	13	9
	p99p99	95	84	40	48	34	24	20	16	10
	maxp99	118	108	49	53	39	26	24	18	11
	p90max	126	116	64	170	207	96	99	59	57
	p99max	239	238	118	199	270	109	127	73	65
	maxmax	297	345	157	221	311	116	163	96	75

[a] For all years monitored, continuous 5-minute measurement concentrations were available.
[b] Adjusted concentrations were based on a monitor-based design value (adjustment factor =55/75 = 0.73).
[c] Adjusted concentrations were based on highest modeled air quality receptor and the primary source contribution to concentrations at that receptor (see section 3.4).
Abbreviations: p90 = 90th percentile of 5-minute concentrations at monitor. p90p90 = 90th percentile of the distribution of all study area receptor 90th percentile 5-minute concentrations

3.5.3.3 Estimated Peak 5-minute Concentrations at Air Quality Receptor Sites During Times of Greater Exposure Potential

Similar to the evaluation conducted on the hourly concentrations (section 3.2.5 and Appendix K), we were interested in understanding how well the estimated 5-minute concentrations corresponded with the available ambient measurements, while focusing on times most likely associated with population exposure and considering all modeled receptors and the estimated 5-minute concentrations used as input to the exposure model. Accordingly, we stratified both the estimated and measurement data sets by time-of-day and season and have focused on the daytime hours during the three warmer seasons.[39]

This analysis used the set of estimated 5-minute concentrations for all air quality receptors in each study area and compared these with the available monitor data, with both sets based on an adjustment intended to reflect the hourly concentrations just meeting the current standard. The 5-minute estimated concentrations for the air quality receptor sites are derived from the hourly concentrations estimated for the current standard scenario (adjustment for hourly concentrations described in section 3.4). As that same approach (which is based on adjusting concentrations associated with emissions from the primary source) could not be applied for the monitor concentrations, monitor concentrations were adjusted by a different approach. Concentrations at the highest monitor in each study area were adjusted such that the 3-year DV just equaled the level of the standard (75 ppb). For areas with more than one monitor, the concentrations at all monitors were adjusted by the same factor (based on the DV monitor). This difference in deriving somewhat conceptually comparable datasets affects our ability to precisely judge the implications of these comparisons with regard to potential bias in the receptor estimates and limits our conclusions accordingly.

Calculated for each data set were instances where 5-minute concentrations were at or above 100, 200, 300, and 400 ppb, at each individual air quality receptor and for each year. These counts developed for each air quality receptor location were then binned using the number of days per year, i.e., a receptor had at least 1 day, 2 or more days, 5 or more days, and 10 or more days at or above a selected level. Then the number of air quality receptor locations in each bin was summed, indicating how many air quality receptor locations in a study area had estimated concentrations at or above the levels of interest. Then we calculated the percentages these numbers were of the total number of receptor sites in each study area. Similar counts and percentages were also calculated for the monitor data. Results generated for each of the three study areas are provided in Table 3-14 to Table 3-16.

In general, there is consistency between the estimated and measured concentrations regarding the number of days per year that concentrations are at or above 5-minute concentrations of interest considering the years and seasons simulated. For example, in Fall River, 6 of the 9 season/years had at least one day with a concentration at or above 100 ppb at the ambient air monitor, while 5 of the 9 season/years were above the same level for more than 40% of air quality receptors in the study area (Table 3-14). The occurrence of these upper percentile concentrations seems more frequent than would be observed when considering the

[39] Data were stratified by two times of day (daytime and nighttime) and four seasons (winter, spring, summer and fall) as described in section 3.2.5.

ambient air monitor alone, particularly for the Indianapolis study area (Table 3-15). Again, this could be a function of the model representing variability in ambient concentrations not observed at the ambient air monitors due to the siting of many modeled receptors in close proximity to the important emission sources in each study area. Comparisons of the estimated and monitor 5-minute concentrations in the Tulsa study area (Table 3-16) were similar to that observed for the other two study areas, although they differed by having a much smaller percent of receptors at or above the concentrations of interest.

Table 3-14. Percent of air quality receptors and monitors at which 5-minute SO_2 concentrations (for conditions just meeting standard) exceed concentrations of interest on single and multiple days, Fall River study area 2011-2013.

5-minute Concentrations of Interest [a]	Season [b]	Year	Percent of Receptors Exceeding Concentration of Interest on Specified Number of Days in Year [c] Number of Days ≥1	≥2	≥5	≥10	Percent of Monitors Exceeding Concentration of Interest on Specified Number of Days in Year [c] Number of Days ≥1	≥2	≥5	≥10
100	Fall	2011	41.7	7.0	0	0	100	100	0	0
		2012	1.6	0.1	0	0	100	100	0	0
		2013	2.3	0.1	0	0	0	0	0	0
	Spring	2011	86.3	60.0	14.1	1.3	100	100	0	0
		2012	0.9	0	0	0	100	100	0	0
		2013	3.9	0.7	0	0	100	0	0	0
	Summer	2011	100	99.9	71.2	12.2	100	100	100	0
		2012	100	100	26.2	4.8	0	0	0	0
		2013	100	100	7.0	0.1	0	0	0	0
200	Fall	2011	0	0	0	0	0	0	0	0
		2012	0	0	0	0	0	0	0	0
		2013	0	0	0	0	0	0	0	0
	Spring	2011	1.3	0.3	0	0	100	0	0	0
		2012	0	0	0	0	0	0	0	0
		2013	0	0	0	0	0	0	0	0
	Summer	2011	1.1	0.8	0	0	100	100	0	0
		2012	0.1	0	0	0	0	0	0	0
		2013	0	0	0	0	0	0	0	0

[a] There were no estimated or measured concentrations at or above 300 ppb.
[b] Daytime hours (6 AM to 8 PM) only.
[c] There were 1,494 receptors modeled and 1 monitor.

Table 3-15. Percent of air quality receptors and monitors at which 5-minute SO_2 concentrations (for conditions just meeting standard) exceed concentrations of interest on single and multiple days, Indianapolis study area 2011-2013.

5-minute Concentrations of Interest	Season [a]	Year	Percent of Receptors Exceeding Concentration of Interest on Specified Number of Days in Year [b]				Percent of Monitors Exceeding Concentration of Interest on Specified Number of Days in Year [b]			
			Number of Days				Number of Days			
			≥1	≥2	≥5	≥10	≥1	≥2	≥5	≥10
100	Fall	2011	100	99.7	60.6	60.5	0	0	0	0
		2012	100	100	63.5	60.5	100	67	33	0
		2013	100	100	60.5	60.5	67	33	33	0
	Spring	2011	60.4	57.3	25.4	2.0	33	33	0	0
		2012	54.6	38.4	5.2	0.7	67	33	0	0
		2013	36.9	16.0	3.0	0.5	33	33	0	0
	Summer	2011	100	100	94.3	60.2	33	33	0	0
		2012	100	100	62.1	60.5	67	33	0	0
		2013	61.0	60.5	60.5	60.5	67	33	0	0
200	Fall	2011	59.5	57.1	0	0	0	0	0	0
		2012	0	0	0	0	33	0	0	0
		2013	60.4	58.2	0	0	0	0	0	0
	Spring	2011	1.3	0.1	0	0	0	0	0	0
		2012	0	0	0	0	33	0	0	0
		2013	0.9	0	0	0	0	0	0	0
	Summer	2011	5.4	1.6	0	0	0	0	0	0
		2012	0.3	0	0	0	33	0	0	0
		2013	25.2	0.6	0	0	33	0	0	0
300	Fall	2011	57.3	0.1	0	0	0	0	0	0
		2012	0.1	0	0	0	0	0	0	0
		2013	0.2	0	0	0	0	0	0	0
	Spring	2011	0.5	0.1	0	0	0	0	0	0
		2012	0.2	0	0	0	33	0	0	0
		2013	0.1	0	0	0	0	0	0	0
	Summer	2011	1.9	0.1	0	0	0	0	0	0
		2012	0.1	0.1	0	0	33	0	0	0
		2013	0	0	0	0	0	0	0	0
400	Fall	2011	2.8	0	0	0	0	0	0	0
		2012	0	0	0	0	0	0	0	0
		2013	0	0	0	0	0	0	0	0
	Spring	2011	0.5	0	0	0	0	0	0	0
		2012	0	0	0	0	0	0	0	0
		2013	0	0	0	0	0	0	0	0
	Summer	2011	0.9	0	0	0	0	0	0	0
		2012	0	0	0	0	0	0	0	0
		2013	0	0	0	0	0	0	0	0

[a] Daytime hours (6 AM to 8 PM) only.
[b] There were 1,917 receptors modeled and 3 monitors.

Table 3-16. Percent of air quality receptors and monitors at which 5-minute SO_2 concentrations (for conditions just meeting standard) exceed concentrations of interest on single and multiple days, Tulsa study area 2011-2013.

5-minute Concentrations of Interest [a]	Season [b]	Year	Percent of Receptors Exceeding Concentration of Interest on Specified Number of Days in Year [c]				Percent of Monitors Exceeding Concentration of Interest on Specified Number of Days in Year [c]			
			Number of Days				Number of Days			
			≥1	≥2	≥5	≥10	≥1	≥2	≥5	≥10
100	Fall	2011	8.2	3.5	2.0	0.7	67	33	0	0
		2012	8.0	2.8	0.9	0.6	33	0	0	0
		2013	0.6	0.4	0.1	0	33	0	0	0
	Spring	2011	5.3	2.1	1.2	0.8	33	33	33	0
		2012	4.5	1.3	0.6	0.6	33	33	33	0
		2013	0.6	0.4	0.1	0	67	33	0	0
	Summer	2011	3.5	1.4	0.9	0.6	33	33	33	33
		2012	2.7	1.0	0.6	0.6	67	33	33	0
		2013	0.6	0.5	0.2	0.1	33	33	0	0
200	Fall	2011	0.6	0.4	0	0	33	0	0	0
		2012	0.5	0.4	0.1	0	0	0	0	0
		2013	0	0	0	0	0	0	0	0
	Spring	2011	0.6	0.4	0.1	0	0	0	0	0
		2012	0.7	0.4	0.3	0.2	33	0	0	0
		2013	0	0	0	0	0	0	0	0
	Summer	2011	0.6	0.4	0.1	0.1	0	0	0	0
		2012	0.5	0.5	0.2	0.1	0	0	0	0
		2013	0	0	0	0	0	0	0	0

[a] There were no estimated or measured concentrations at or above 300 ppb.
[b] Daytime hours (6 AM to 8PM) only.
[c] There were 1,389 receptors modeled and 3 monitors.

REFERENCES

Casella G and Berger RL. (2002). Statistical Inference. Second Edition. Editor: Carolyn Crocket, Duxbury/Wadsworth Group, Pacific Grove CA.

Junger WL and de Leon AP. (2015). Imputation of missing data in time series for air pollutants. Atmospheric Environment. 102:96-104.

Kahn HD. (1973). Note on the distribution of air pollutants. Journal of the Air Pollution Control Association. 23(11):973.

Karl TR and Koss, WJ. (1984). Regional and National Monthly, Seasonal, and Annual Temperature Weighted by Area, 1895-1983. Historical Climatology Series 4-3, National Climatic Data Center, Asheville, NC, 38 pp. Available at: https://www.ncdc.noaa.gov/monitoring-references/maps/us-climate-regions.php

Larsen R. (1977). An air quality data analysis system for interrelating effects, standards, and needed source reductions: Part 4. a three-parameter averaging-time model. Journal of the Air Pollution Control Association. 27(5):454-459.

Rizzo M. (2009). Memorandum to the Sulfur Dioxide Review Docket (EPA-HQ-OAR-2007-0352). Investigation of How Distributions of Hourly Sulfur Dioxide Concentrations Have Changed Over Time in Six Cities. Sulfur Dioxide Review Docket. Docket ID No. EPA-HQ-OAR-2007-0352-0017.

SAS. (2017). Base SAS® 9.4 Procedures Guide, Seventh Edition. Available at: http://documentation.sas.com/api/collections/pgmmvacdc/9.4/docsets/proc/content/proc.pdf?locale=en#nameddest=bookinfo. Also used were SAS/ETS 14.2 User's Guide, edited by A. Baxter, E. Huddleston and SAS/QC 14.2 User's Guide, edited by A. Baxter, V. Clark, E. Huddleston, S. Prabhu, R. Rodriguez, D. Sawyer, J. Simmons.

Singer I. (1961). The relationship between peak and mean concentrations. Journal of the Air Pollution Control Association. 11(7):336-341.

Turner B. (1964). A diffusion model for an urban area. Journal of Applied Meteorology. 3(1):83-91.

U.S. EPA. (2009). Risk and Exposure Assessment to Support the Review of the SO_2 Primary National Ambient Air Quality Standard. Office of Air Quality Planning and Standards, Research Triangle Park, NC, EPA-452/R-09-007, July 2009. Available at: https://www3.epa.gov/ttn/naaqs/standards/so2/data/200908SO2REAFinalReport.pdf

U.S. EPA. (2013). AERSURFACE User’s Guide. Office of Air Quality Planning and Standards, Research Triangle Park, NC, EPA-454/B-08-001, Revised January 2013. Available at: https://www3.epa.gov/scram001/7thconf/aermod/aersurface_userguide.pdf

U.S. EPA. (2015a). AERMINUTE User's Guide. Office of Air Quality Planning and Standards, Research Triangle Park, NC, EPA-454/B-15-006, October 2015. Available at: https://www3.epa.gov/ttn/scram/7thconf/aermod/aerminute_userguide.pdf

U.S. EPA. (2015b). 2011 National Emissions Inventory (NEI) Technical Support Document, version 2. Office of Air Quality Planning and Standards, Research Triangle Park, NC, August 2015. Available at: https://www.epa.gov/sites/production/files/2015-10/documents/nei2011v2_tsd_14aug2015.pdf

U.S. EPA. (2016a). Guideline on Air Quality Models. 40 CFR Part 51 Appendix W.

U.S. EPA. (2016b). User's Guide for the AMS/EPA Regulatory Model – AERMOD. EPA-454/B-16-011. Office of Air Quality Planning and Standards, Research Triangle Park, NC.

U.S. EPA. (2016c). User's Guide for the AERMOD Meteorological Processor (AERMET). Office of Air Quality Planning and Standards, Research Triangle Park, NC, EPA-454/B-16-010.

U.S. EPA. (2016d). User's Guide for the AERMOD Terrain Preprocessor (AERMAP). Office of Air Quality Planning and Standards, Research Triangle Park, NC, EPA-454/B-16-012, December 2016.

U.S. EPA. (2016e). Technical Support Document (TSD): Preparation of Emissions Inventories for the Version 6.3, 2011 Emissions Modeling Platform. Office of Air Quality Planning and Standards, Research Triangle Park, NC, August 2016. Available at: https://www.epa.gov/sites/production/files/2016-09/documents/2011v6_3_2017_emismod_tsd_aug2016_final.pdf

U.S. EPA. (2016f). SO_2 NAAQS Designations Modeling Technical Assistance Document, Draft. Office of Air Quality Planning and Standards, Research Triangle Park, NC, August 2016. Available at: https://www.epa.gov/sites/production/files/2016-06/documents/so2modelingtad.pdf

U.S. EPA. (2016g). AERMOD Implementation Guide. Office of Air Quality Planning and Standards, Research Triangle Park, NC, EPA-454/B-16-013, December 2016.

U.S. EPA. (2017a). Air Markets Program Data. Available at: https://ampd.epa.gov/ampd/

U.S. EPA. (2017b). Review of the Primary National Ambient Air Quality Standard for Sulfur Oxides: Risk and Exposure Assessment Planning Document. Office of Air Quality Planning and Standards, Research Triangle Park, NC, EPA-452/P-17-001, February 2017. Available at: https://www3.epa.gov/ttn/naaqs/standards/so2/data/20170216so2rea.pdf

4 POPULATION EXPOSURE AND RISK

This chapter describes the methods used to characterize exposure and health risk associated with SO_2 emitted into ambient air under conditions just meeting the current primary standard. As summarized in section 2.2, the overall analysis approach is based on linking the health effects information to estimated population-based exposures that reflect our current understanding of 5-minute concentrations of SO_2 in the ambient air.

Population exposures were estimated using the EPA's Air Pollution Exposure Model (APEX), version 5. The APEX model is a multipollutant, population-based, stochastic, microenvironmental model that can be used to estimate human exposure via inhalation for criteria and toxic air pollutants. APEX is designed to estimate human exposure to these pollutants at the local, urban, and consolidated metropolitan level. In this REA we have used APEX to estimate exposures in three study areas, the details of which are provided in the following subsections. Additional information not provided here regarding all of APEX modules, algorithms, and model options can be found in the APEX User's Guide (U.S. EPA, 2017a, b).

Briefly, APEX calculates the exposure time-series for a user-specified exposure duration and number of individuals. Collectively, these simulated individuals are intended to be a representative random sample of the population in a given study area. To this end, demographic data from the decennial census are used so that appropriate model sampling probabilities can be derived, considering personal attributes such as age and sex, and used to properly weigh the distribution of individuals in any given geographical area. For this REA, the core demographic geographical units for estimating exposure are census blocks. Because AERMOD predicted SO_2 concentrations at air quality receptor sites in a regular spaced grid (chapter 3), APEX matches the centroid of each census block (which are irregularly spaced due to varying size) in the study area with the closest receptor to estimate exposures for simulated individuals residing in each census block.

For each simulated person, the following general steps are performed:

- Select attribute variables and choose values to characterize the person (e.g. age, sex, body weight, disease status);
- Construct the activity event sequence (minute by minute time series) by selecting a sequence of appropriate daily activity diaries for the person (using demographic and other influential variables);
- Calculate the concentrations in the microenvironments (MEs) that simulated individuals visit;
- Calculate the person's simultaneous breathing rate and exposure for each event and summarize for the selected exposure metric.

A simulated individual's complete time-series of exposures (i.e., *exposure profile*), representing intra-individual variability in exposures, is combined with the exposure profiles for all simulated individuals in each study area and summarized to generate the population distribution of exposures, representing inter-individual variability in exposures. As described above regarding air quality and in the sections that follow describing APEX model inputs and approaches to estimating exposure, the overarching goal of the REA is to account for the most significant factors contributing to inhalation exposure, i.e., the temporal and spatial distribution of people and pollutant concentrations throughout the study area and among the microenvironments. The population distributions of exposures are combined with the health effects information to characterize associated risk via two types of metrics: comparison to benchmark concentrations and lung function risk. The details of the methods for exposure and risk estimation are described in the sections that follow.

4.1 POPULATIONS SIMULATED

APEX stochastically generates a user-specified number of simulated persons to represent the population in the study area. The number of simulated individuals can vary and is dependent on the size of the population to be represented. In these analyses, the number of simulated individuals was set at 100,000 in each area, a more than adequate number of individuals to represent the geographically-restricted population residing within the exposure modeling domains (approximately 180,000 – 500,000). Each simulated person is represented by a "personal profile." The personal profile includes characteristics such as a specific age, a specific home sector, a specific work sector (or does not work), specific housing characteristics, specific physiological parameters, and so on. The profile does not correspond to any particular individual in the study area, but rather represents a simulated person. Accordingly, while a single profile does not, in isolation, provide information about the study population, a distribution of profiles represents a random sample drawn from the study area population. This means that the modeling objective is for the statistical properties of the distribution of profiles to reflect statistical properties of the population in the study area.

APEX generates population-based exposures using several population databases. Based on the geographic boundaries defining the study areas and the study groups of interest, APEX will simulate representative individuals using appropriate geographic, demographic, and health status information provided by existing population-based surveys. In this REA, there is variation in the geographic units by which some of the input data sets are organized (e.g., U.S. census tracts or a smaller subdivision such as census blocks). For example, employment status data are provided at the tract level while population demographics are available at the block level. Regardless of the geographic unit of the input data, all population-based data sets were applied at the block level in the exposure simulations. Where only tract level data were available, we assigned the tract specific information directly to the blocks that comprise a particular tract.

Several updates were made to the APEX model inputs and algorithms for use in simulating the populations of interest in this REA and are described in the following sections: (1) population demographic data that are based on the 2010 census (section 4.1.1), (2) asthma prevalence rates based on the 2011-2015 National Health and Nutrition Examination Survey (NHANES) and vary by age, sex and geographic location (section 4.1.2), and data and equations used to approximate personal attributes such as body weight, resting metabolic rate, and breathing rate (section 4.1.3).

4.1.1 Demographics

As described in section 3.2.3, ambient air concentrations were modeled to a fine-scale grid (100 m – 2 km) in each study area to better capture spatial heterogeneity in ambient air SO_2 concentrations. We used U.S. Census blocks, the finest geographical scale available for the population data,[1] to take full advantage of this fine-scale air quality surface and best match the potential at-risk populations with areas having the highest SO_2 concentrations. Block-level population counts were obtained from the 2010 Census of Population and Housing Summary File 1.[2] Summary Files 1 (SF1) contains what the Census program calls "the 100-percent data," which is the information compiled from the questions asked of all (100% of) people and housing units in the U.S. Three standard APEX input files[3] are used for the current assessment. For the purposes of having a more tractable analysis, we restricted these population demographic files to include the census blocks within the five states that encompass the three study areas (i.e., Connecticut, Indiana, Massachusetts, Oklahoma, and Rhode Island) rather than use a national-based file that would include all 50 U.S. states.

- *PopGeoLocs2010_3StudyAreas.txt*: census block identifiers (ID's), latitudes and longitudes in degrees.

[1] The minimum size for census block is between 30,000 to 40,000 ft^2 or approximating a grid cell of about 55 – 60 meters (see https://www.census.gov/geo/reference/garm.html). The next larger sized census geographic unit is a census block group which is comprised of multiple blocks. When considering that, on average, there are about 30 to 85 blocks per block group in the states where the study areas are located, it is likely that census block groups would be more amenable to a modeled air quality surface having a grid cell size of about 1.6 – 8.1 Km (see https://www.census.gov/geo/maps-data/data/tallies/tractblock.html).

[2] Technical documentation - 2010 Census Summary File 1—Technical Documentation/prepared by the U.S. Census Bureau, Revised 2012 - available at: http://www.census.gov/prod/cen2010/doc/sf1.pdf.

[3] The names of all APEX files are provided here to link the brief description with the appropriate input file.

- *PopBlockFemale2010_3StudyAreas.txt*: census block identifiers, block-level population counts for females, stratified by 23 age groups.[4]
- *PopBlockMale2010_3StudyAreas.txt*: census block identifiers, block-level population counts for males, stratified by 23 age groups.

We evaluated the spatial distribution of the population in each study area, focusing on children, a study group identified as an important at-risk population (section 2.1.3). First, we subset this APEX input data set to include the blocks that are a part of the study areas. Then, because there are wide ranging numbers of people in each block, we stratified these data into three population groups: blocks having at least 100 people, blocks having greater than 100 to at least 500 people, and blocks having greater than 500 people. Finally, we calculated the percent of the total population that were children (aged 0-17) in each census block of each study area and population group, and calculated the percentiles of that distribution providing perspective on the spatial distribution of children (and adults) in each study area (Table 4-1).

In each study area there are a number of blocks having no people residing in them (i.e., non-residential blocks). While these blocks are retained in the exposure simulations as they could still serve as an area where an individual might visit and be exposed to SO_2 (e.g., a workplace location within a study area), these blocks were not used to calculate the population spatial distribution statistics. The majority of the residential blocks (84-93%) in each study area have fewer than 100 people, with very few blocks (<1%) having a total population greater than 500 people. From a relative perspective, the Indianapolis study area had approximately double the percent of residential blocks having a total population greater than 500 people compared to the other two study areas. Further, while the overall distribution of the percent of children in each study area is general similar (comprising 15-20% on average per residential block), the Indianapolis study area consistently has a greater percent of children across most of the percentiles of the distribution, most notably so for the residential blocks with a population greater than 500 people.

[4] The age groups in this file are: 0-4, 5-9, 10-14, 15-17, 18-19, 20-20, 21-21, 22-24, 25-29, 30-34, 35-39, 40-44, 45-49, 50-54, 55-59, 60-61, 62-64, 65-66, 67-69, 70-74, 75-79, 80-84, >84.

Table 4-1. Distribution of the percent of total population that are children residing in the census blocks comprising each study area.

	Percent of Population that are Children Residing in Study Area Census Blocks								
	Census blocks with ≤100 people per block			Census blocks with >100 to ≤500 people per block			Census blocks with >500 people per block		
Study area [a]	**FR**	**IN**	**OK**	**FR**	**IN**	**OK**	**FR**	**IN**	**OK**
Mean (%)	14.7	16.9	16.7	16.4	19.7	17.7	14.6	18.4	11.9
Standard deviation (%)	10.0	11.0	11.2	7.2	9.0	9.8	8.2	8.9	8.7
Minimum (%)	0	0	0	0	0	0	0.5	0	0.5
5th percentile (%)	0	0	0	4.3	2.6	0.9	0.5	0.2	0.65
10th percentile (%)	0	0	0	8.3	7.8	3.8	3.05	2.8	1.2
25th percentile (%)	7.75	9.1	8.8	12.3	14.5	11.4	8.9	14	3
50th percentile (%)	14.3	17	16.7	16.1	19.7	18	14.8	19.5	11.1
75th percentile (%)	20.3	24	23.9	19.9	25.4	23.8	22	25.8	17.7
90th percentile (%)	27.3	30.4	30.9	24.5	30.5	30	23.9	28.5	24.95
95th percentile (%)	31.9	34.3	35.1	28.2	33.9	34.9	25.3	30.6	27.9
99th percentile (%)	42.3	46.2	47.5	39.4	42.3	41.7	25.3	34.5	28.5
Maximum (%)	86.2	100	83.3	51.7	61.6	61.3	25.3	34.5	28.5
Number of blocks	3883	11130	7319	471	1123	355	10	57	20
Number of blocks with non-zero population	2588	7650	5106	471	1123	355	10	57	20

[a] FR = Fall River, IN = Indianapolis, OK = Tulsa.

The employment file for APEX contains the probability of employment separately for males and females by age group (starting at age 16) and by Census tract (the only census unit available for this type of data). The 2010 Census collected basic population counts and other data using the short form, but collected more detailed socioeconomic data (including employed persons) from a relatively small subset of people using the 5-year American Community Survey (ACS).[5] The ACS dataset provides the number of people in the labor force, which we stratified by sex/age/tract, considering both civilian workers and workers in the Armed Forces. The data were stratified by sex and age group, and were processed so that each sex-age group combination is given an employment probability fraction (ranging from 0 to 1) within each census tract. Children under 16 years of age were assumed to be unemployed. To match the population

[5] 2010 U.S. Census American FactFinder: http://factfinder2.census.gov/. For instance, to obtain the table ID B23001 "Sex by age by employment status for the population 16 years and over", the following steps were performed. First, select the "guided search option", choose "information about people" and select "employment (labor force) status", "sex" and "age". For geography type select "census tract - 140" for each state. Tables containing the employment numbers were downloaded and used to calculate the employment probabilities for each age group.

demographic files, we included only the census blocks within the five states that encompass the three study areas. To use the file at a block level, all blocks were assigned the same employment probabilities as the parent tract. One standard APEX input file is used for the current assessment:

- *EmpBlock2010_3StudyAreas.txt*: census block IDs, employment probabilities (in fractional form), stratified by 13 age groups.[6]

4.1.2 Asthma Prevalence

The population groups included in this exposure assessment are adults with asthma (> 18 years old) and children with asthma (5 to 18 years old),[7] based on their identification in this REA as an at-risk population (section 2.1.3). To best approximate the number (and percent) of individuals comprising each of these population groups in each study area, we considered several influential variables that could affect asthma prevalence. It is widely recognized that there are significant differences in asthma prevalence based on age, sex, U.S. region, and family income level, among other factors.[8] There is spatial heterogeneity in family income level across census geographic areas (and also across age groups)[9] and spatial variability in local scale ambient air concentrations of SO_2 (e.g., Figures 3-6 to 3-8). Thus, we have developed an approach to better estimate the variability in population-based SO_2 exposures by accounting for these particular attributes of this study group and their spatial distribution across each of the study areas.

With regard to asthma prevalence, the data are used to identify if a simulated individual residing within a modeled census geographic area has asthma – and are not used for selection of any other personal attribute nor in the selection of activity pattern data. Thus, our primary objective with these data was to generate census block level prevalence estimates that reflect variability in asthma prevalence contributed by several known influential attributes (e.g., age, sex, geographic location). Two data sets were identified and linked together to estimate asthma prevalence used for this REA. First, asthma prevalence data were obtained from the 2011-2015 National Health Interview Survey (NHIS) and are stratified by NHIS defined regions (Midwest, Northeast, South, and West), age, and sex.[10] We explored other variables that were available in

[6] The age groups in this file are: 16-19, 20-21, 22-24, 25-29, 30-34, 35-44, 45-54, 55-59, 60-61, 62-64, 65-69, 70-74, and >75.

[7] As in other NAAQS reviews, this REA does not estimate exposures and risk for children younger than 5 years old due to the more limited information contributing relatively greater uncertainty in modeling their activity patterns and physiological processes than children between the ages of 5 to 18.

[8] For example, see the Center for Disease Control report “National Surveillance of Asthma: United States, 2001–2010”, available at: https://www.cdc.gov/nchs/data/series/sr_03/sr03_035.pdf.

[9] For example, see the U.S. Census report “Income and Poverty in the United States: 2016”, available at: https://www.census.gov/content/dam/Census/library/publications/2017/demo/P60-259.pdf.

[10] Information about the NHIS is available at: http://www.cdc.gov/nchs/nhis.htm.

the NHIS data set that contributed to variability in asthma prevalence and that could be used to extrapolate the asthma prevalence to a finer geographic scale than the NHIS-provided four regions. The linking variable had to be common with variables available in the population demographic data. Based on this criterion, we selected family income level to poverty thresholds (i.e., whether the family income was considered below or at/above the US Census estimate of poverty level for the given year) and used that as an additional variable to stratify the NHIS asthma prevalence. Then, we obtained information from the 2013 Census ACS to estimate family income level to poverty thresholds at the census tract level and stratified by several ages and age groups.[11] By combining these two data sets, we developed census tract level asthma prevalence estimates for children (by age in years) and adults (by age groups), also stratified by sex (male, female) that were weighted by the individual census tract populations and family income level. To match the population demographic and employment files, we included only the census blocks within the same five states that encompass the three study areas. The census tract sex- and age-specific prevalence were extrapolated to census blocks using the 11-character identifier shared between census tracts and blocks. A detailed description of how the NHIS data were processed to create the data set used for input to APEX is provided in Appendix E. One standard APEX input file is used for the current SO_2 assessment:

- *asthma_prev_1115_block_3StudyAreas.txt*: census block identifiers, block-level asthma prevalence (in fractional form) stratified by sex, 18 single year ages (for ages <18),[12] and 7 age groups (for ages > 17).

The asthma prevalence estimates vary for the different ages and sexes of children and adults[13] that reside in each census block of each study area. We evaluated the spatial distribution of the asthma prevalence using the specific blocks that comprise the exposure modeling domain in each study area. We first separated the estimates for children from those for adults and calculated the distribution of asthma prevalence for the blocks, stratified by sex (Table 4-2). By design (i.e., the use of age, sex, and family income variables), there is spatial variability in the

[11] Census tract level data is the finest scale geographical unit having family income information. The family income/poverty ratio threshold used was 1.5, that is the surveyed person's family income was considered either ≤ or > than a factor of 1.5 of the U.S. Census estimate of poverty level for the given year.

[12] The census data set used only had children for single years up to and including age 17, after that year they are provided in groups. The upper portion of this age range differs from those considered as children in estimating exposures i.e., in our exposure assessment children are considered upwards to 18 years old. To simulate the number of children with asthma age 18, estimated prevalence from the first adult group were used (i.e., individuals age 18-24).

[13] While prevalence rates were estimated for all ages of children (in single years 5 - 17), for adults they were estimated for seven age groups: 18-24 years, 25-34 years, 35-44 years, 45-54 years, 55-64 years, 65-74 years, and, ≥75 years old (see Appendix E for more information).

prevalence estimates. Consistent with broadly defined national asthma prevalence (e.g., Table 3-2 of SO_2 PA), children have higher estimated rates than adults,[14] male children have higher rates than female children,[15] and adult females have higher rates than adult males (e.g., compare with mean values of Table 4-2). However, when evaluating variability contributed by study area, age, sex, and family income level on a spatial scale, an additional degree of variability emerges across the study areas (as presented in Tables 4-2). The Fall River study area has some of the highest asthma prevalence for children considering most of the statistics with rates as high as 21.5% in one or more census blocks for males of a given year of age. The Tulsa study area exhibits some of the lowest asthma prevalence when considering adults (both sexes) with rates as low as 4.0% in one or more blocks for males within a given age group. These summary statistics represent the range of age- and sex-specific values for the census blocks used in each APEX simulation to estimate the number of individuals that have asthma.

Table 4-2. Estimated asthma prevalence for children and adults in census blocks of three study areas, summary statistics.

Study Area (# census blocks) and **Population group**		**Sex**	**Asthma Prevalence across all ages (or age groups) for all census blocks** [a]				
			Mean	**Standard Deviation**	**Minimum**	**Median**	**Maximum**
Fall River (4,364)	child	female	9.3%	2.4%	5.7%	9.2%	18.6%
		male	13.3%	2.3%	8.4%	13.3%	21.5%
	adult	female	9.9%	1.5%	7.2%	9.8%	17.6%
		male	6.3%	1.0%	5.1%	5.8%	9.0%
Indianapolis (12,310)	child	female	9.1%	2.0%	5.8%	8.6%	19.4%
		male	10.8%	2.3%	6.6%	10.7%	16.8%
	adult	female	9.9%	1.8%	6.8%	10.0%	17.6%
		male	6.1%	1.5%	2.5%	5.9%	10.4%
Tulsa (7,694)	child	female	10.2%	1.5%	7.3%	10.2%	13.9%
		male	12.0%	2.0%	7.5%	12.3%	16.1%
	adult	female	8.8%	1.7%	5.5%	8.8%	14.4%
		male	5.1%	0.6%	4.0%	5.0%	6.9%

[a] As described in text above this table, prevalence estimates are based on age (or age group) and sex-specific prevalence estimates for each census block derived from CDC NHIS asthma prevalence and U.S. census income/poverty ratio information.

[14] Asthma prevalence, when not separated by sex, is greater for children (mean of 11.3%, 10.0%, and 11.1%) than that of adults (mean of 8.1%, 8.0%, 7.0%) for the Fall River, Indianapolis, and Tulsa study areas, respectively. Nationally, asthma prevalence for children is 8.4% and for adults is 7.6% (Table 3-2 of SO_2 PA).

[15] Asthma prevalence, when not separated by the three study areas evaluated, is greater in boys (mean of 11.6%) than that of girls (mean of 9.4%). Nationally, asthma prevalence for boys is 9.9%, for girls is 6.9% (Table 3-2 of SO_2 PA).

There are many other personal attributes that have been shown to influence asthma prevalence, such as race, ethnicity, obesity, smoking, health insurance, and activity level (e.g., Zahran and Bailey, 2013). The set of variables chosen to stratify asthma prevalence for use in this REA (i.e., age, sex, and family income level) was based on 1) maximizing the potential range in asthma prevalence variability, 2) maximizing the number of survey respondents comprising a representative subset study group, and 3) having the ability to link the set of attributes to variables within the Census population demographic data sets. Many of the additional potential influential factors identified here are not available in the census data and/or have limited representation in the asthma prevalence data (e.g., the survey participant has health insurance or they provide a response to a question regarding their body weight). Race is perhaps the only attribute common to both the prevalence and population data sets that could be an important influential factor and was not directly used in this REA to calculate asthma prevalence. However, the use of race in calculating asthma prevalence, either alone or in combination with family income level, would further stratify the NHIS analytical data set and appreciably reduce the number of individuals of specific age, sex, race, and family income level, potentially reducing the confidence in calculated asthma prevalence based on so few data. Because family income level already strongly influences asthma prevalence across all races and stratifies the NHIS data into only two subgroups (i.e., above or below the poverty threshold) rather than the larger number of subgroups a race variable might yield, family income was chosen as the next most important variable beyond age and sex to rely on for weighting the spatial distribution of asthma prevalence.

That said, there is some uncertainty in our estimates caused by not utilizing race as a influential variable to spatially weight asthma prevalence (e.g., in addition to family income level). Therefore, we evaluated the influence race and obesity (as indicated by body mass index or BMI) might have on asthma prevalence.[16] While considering this, we also note that while census demographic data are available on the spatial distribution of variables such as race, age, sex and income level, we are unaware of useful data for spatially allocating obesity prevalence across a geographic area.

We first evaluated the number of NHIS adult and child participants that provided race and BMI information and compared them to the numbers of these groups that provided the information (age, sex, family income) used in the REA approach. The adult data set had few missing values when considering the new variables race and BMI (totaling about 160,000

[16] The processing of the prevalence data is described in Appendix E and considered age, sex, and family income level for each of four regions (i.e., Northeast- NE, Midwest-MW, South-S, and West-W). We evaluated these same variables here, however, we also included data responses from the variable RACERPI2 (a value of "2" represented black African Americans) and the variable BMI (or BMI_SC) for race and obesity, respectively.

adults); however, only one-third of children had values for BMI. Thus, the analytical data for children set was substantially smaller in survey sample size compared to the NHIS data set used for the REA approach (i.e., 20,000 vs 60,000 participants). The smaller sample size of children responding to this survey question indicates a potentially greater source of uncertainty with the use of this data set as compared to the relatively more complete adult data set for these variables.

A logistic model was constructed using PROC SURVEYLOGISTIC (SAS, 2017) and included the newly expanded list of potentially influential variables. Age was considered as a continuous variable while four other variables were evaluated using a dichotomous form: sex (female to not); family-income ratio (below poverty to not); race (black African American to not); BMI (obese or BMI≥ 30 to not). Models were then applied to this data set using the same four regional stratifications. We first evaluated the independent effect each variable had on asthma prevalence (i.e., the individual responded "yes" to the question about still having asthma). Odds ratios were calculated and plotted along with 95% confidence intervals.[17] The odds ratios can be interpreted as the percent difference between the two conditions being compared and the variable's impact on the estimated asthma prevalence. Results of this analysis for each of the variables is presented in Figure 4-1 (adults) and Figure 4-2 (children). We also evaluated all possible variable interactions in the statistical model. While the results are somewhat variable across the data sets and complicated to interpret, in general many of the variable independent effects remained statistically significant, and there were few significant interactions (Appendix E, Attachment 4).

Overall, the results for the adult data were more consistent across the different U.S. regions than the child data, possibly due to having a less complete data set for the children. Obesity, rather than race, appears to play a more important role in influencing the asthma prevalence in adults (Figure 4-1) compared to children, while both obesity and race appear to play an important role in explaining asthma prevalence in children (Figure 4-2). Family income level was important alone and in interactions with other influential variables (e.g., BMI, race) considering both children and adults and for most of the four U.S. regions. Further, family income level consistently exhibited greater influence than race on adult asthma prevalence in three of the four regions, while race exhibited a greater influence than family income when considering child asthma prevalence. Sex also had a greater influence on asthma prevalence in

[17] The odds ratios can be interpreted as the percent difference between the two conditions being compared and the variable's impact to the estimated asthma prevalence. For instance, Figure 4-1 shows that in the Northeast, an odds ratio of 1.85 was calculated for adult asthma prevalence considering the BMI variable (i.e., obese - BMI ≥30 vs. not obese - BMI<30). This suggests that if an individual is obese, they are 85% more likely to have asthma than an individual that is not obese. In addition, the 95% confidence intervals that include a value of 1.00 are not considered statistically significant, thus when considering the BMI variable, statistical significance ($p<0.05$) can be assigned to the effect this variable has on the asthma prevalence.

adults compared to children, though women are more likely to have asthma than men while boys are more likely to have asthma than girls. Age was generally not a large factor as assessed in the constructed model (i.e., the statistical test evaluates year-to-year comparative differences), though clearly asthma varies across the full lifetime of ages (e.g., see Appendix E, Attachments 1 and 2) and is an important variable for spatially linking at risk populations and ambient air concentrations in the exposure model.

This evaluation indicates that use of the family income variable in this exposure assessment as an influential factor to estimate variation in asthma prevalence provides reasonably similar estimates as race, particularly for adults. For children, while family income was shown as an important influential variable and, in a few instances, could serve as a surrogate variable to approximate the degree of race-related influence, family income alone may not entirely explain spatial variability in asthma prevalence across urban areas such as those in this REA.

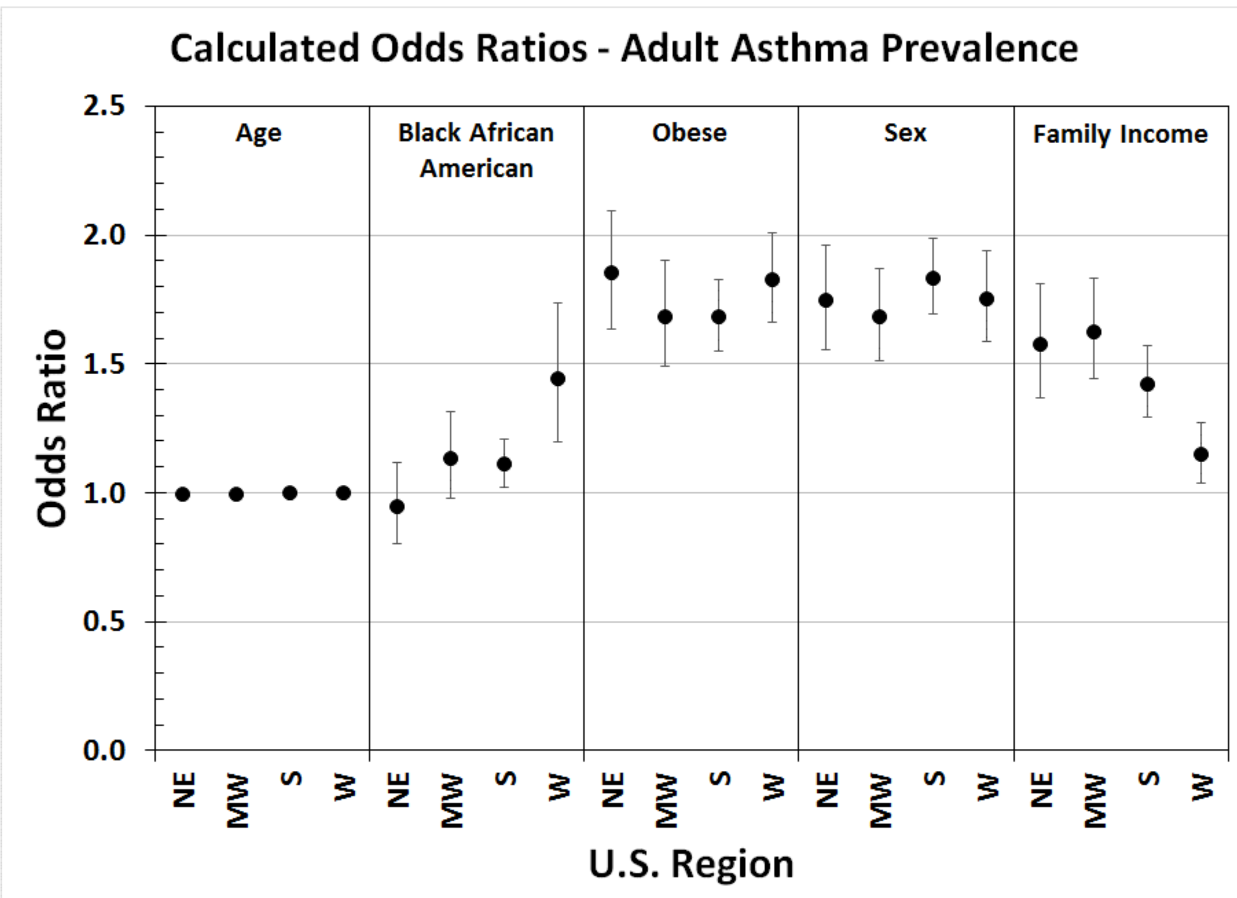


Figure 4-1. Influence of age, race, obesity, sex and family income on adult asthma prevalence (based on NHIS 2011-2015 for four U.S. regions).

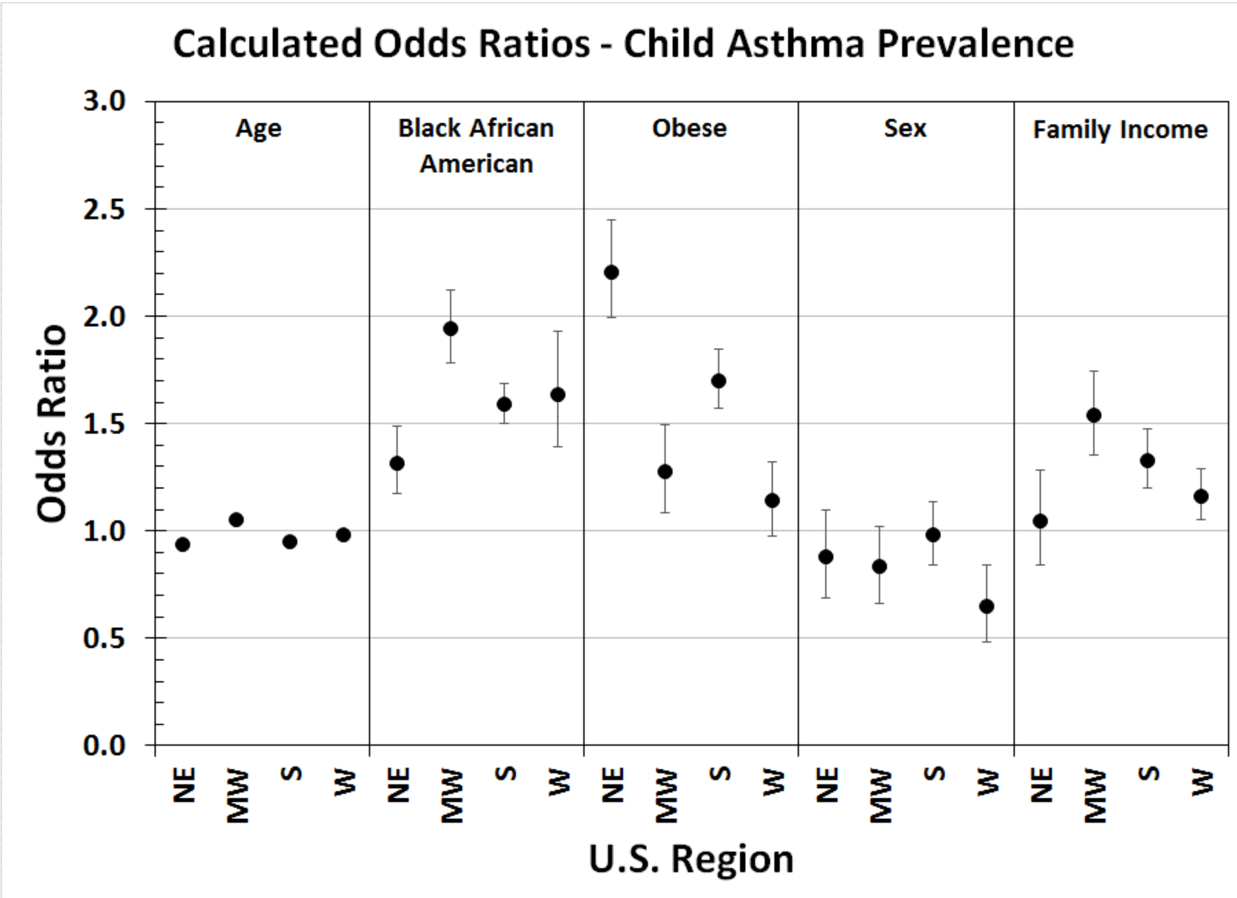


Figure 4-2. Influence of age, race, obesity, sex and family income on child asthma prevalence (based on NHIS 2011-2015 for four U.S. regions).

We note, however, the need for caution in interpreting the influence of race on child asthma prevalence due to uncertainty associated with the sharply reduced dataset for which both BMI and race of the child is specified. It would be possible, however, to generate asthma prevalence using both race and family income as influential variables. If then used in an exposure assessment with the same approach and other inputs in this REA, it is possible that there could be a greater percent and number of people with asthma exposed to concentrations at or above a concentration of interest if both high concentrations and high asthma prevalence coincide spatially to a greater extent than represented here.

Finally, while obesity appears to influence asthma prevalence in both children and adults, currently, it is not a personal attribute that can be used in selecting activity patterns nor do data exist to allow for spatial allocation across a study area. Note also that the body mass (and height if needed) of simulated individuals is estimated by APEX through random sampling of statistical distributions derived from recent NHANES data (section 4.1.4.2), which appropriately reflects an overall distribution of body mass for a given population, and includes simulated individuals that would be characterized as obese. However, the resulting distribution of body mass in the simulated population is not likely to reflect body mass variation at a local-scale in any particular study area.

4.1.3 Personal Attributes

In addition to using the above demographic information to construct the simulated individuals, each modeled person is assigned anthropometric and physiological attributes. All of these variables are treated probabilistically in APEX, taking into account interdependencies where possible, and reflecting variability in the population. It is not the intention of this document to provide detailed description of all the model inputs in each of the files and the data used in their derivation, however there are a few that have been recently updated for use in this REA, namely new statistical distributions for estimating body weight, equations for estimating resting metabolic rate, and equations for estimating activity specific ventilation rate. Brief descriptions of the data used to develop the input files are provided in the sections below. For additional detail, see Appendices F through H and the data within the APEX input files.

4.1.3.1 Body Weight and Surface Area

Anthropometric attributes utilized by APEX in various assessments for estimating pollutant-specific exposures or doses include height, body weight (BW), and body surface area (BSA). Two key personal attributes determined for each individual in this assessment are BW and BSA, both of which are used in the calculation of a number of other variables associated with estimating exposures (e.g., ventilation rate).

Regarding the estimation of body weight, a new APEX input file was generated using 2009-2014 NHANES data.[18] Briefly, body weight and height data for surveyed individuals were obtained and stratified by sex and single years for ages 0 – 79; all ages above 80 were combined as a single age group. Statistical form of the age- and sex-specific body weight and height distributions were evaluated using a log-likelihood statistic. Body weight was found to best fit a lognormal distribution; height was found to best fit a normal distribution. Because height and body weight are not independent, the joint distributions of height and logarithm of body weight were fit assuming a bivariate normal distribution. Then, parameters defining the joint distributions[19] were smoothed using a natural cubic spline to have them represent continuous functions of age rather than vary discontinuously. In addition, having the smoothed parameters

[18] Original data are available in the form of the questionnaire datasets for 2009-2010, 2011-2012, 2013-2014 NHANES from pages reached from this main page: https://wwwn.cdc.gov/nchs/nhanes/Default.aspx. Details regarding the data used and the derivation of the distributions is provided in Appendix G.

[19] Five parameters were used for each age and sex: mean log(BW), standard deviation of log (BW), mean (height), standard deviation of (height), and body weight height correlation coefficients.

could be used to extrapolate information obtained from the single age year distributions (ages 0 – 79) to approximate statistical distributions of body weight for ages ≥ 80. A linear function was fitted to ages 70 and above to extrapolate the parameter values (and hence the statistical distributions of body weight) up to age 100. These distributions are randomly sampled to estimate an age and sex-specific body weight for each simulated individual. Comparison of the new distributions to the body weight distributions previously used by APEX and developed from the 1999-2004 NHIS indicate, for both sexes and across all ages, simulated body weight is about two percent greater using the updated distributions. This difference is expected given the consistent trend of increasing body weight that has occurred over the past few decades.

Finally, age- and sex-specific body surface area, a variable used in conjunction with breathing rate to approximate moderate or greater exertion (section 4.1.3.3) is estimated for each simulated individual as shown in Equation 4-1, and is based on an analysis provided by Burmaster (1998).

$$BSA = e^{\mathbf{-2.2781}} \times BW^{\mathbf{0.6821}} \qquad \text{Equation 4-1}$$

One standard APEX input file is used for the SO_2 assessment:

- *Physiology040617_noHT_Graham_Glen_QA.txt*: Provides parameters for estimating body weight (log BW, standard deviation of BW, lower and upper bounds of BW, by single age years 0-100 and by two sexes) and regression coefficients used in estimating BSA for all sexes and ages.

4.1.3.2 Energy Expenditure and Oxygen Consumption

Energy expended by different individuals engaged in different activities can have an important role in pollutant-specific exposure and/or dose. For example, energy expenditure is related to ventilation rate, which is an important variable in this REA given that the SO_2-induced lung function response has been documented to occur under conditions of elevated ventilation (section 2.1.4 above). In addition, because we are also interested in exposures that occur over short durations (i.e., 5-minutes), estimating activity-specific ventilation rate ($\dot{V}_E$) has been an important motivation behind the development of the algorithm used by APEX. The fundamental basis for $\dot{V}_E$ algorithm is founded in energy expenditure which, for our modeling purposes here, can be related to an individual's resting metabolic rate (RMR) or the energy expended while an individual is at complete rest, along with the energy expended while an individual performs activities involving greater exertion, termed here as metabolic equivalents of work (METs). The approaches used by APEX for estimating RMR and METs are described below, beginning first with the update to the equations used for estimating an individual's RMR.

To estimate RMR for the 2009 REA, the previous version of APEX (version 4.3) had used an algorithm originally based on analyses by Schofield (1985). Because all of the clinical subject data used by Schofield (1985) were from studies conducted as far back as 60 years prior to that publication, we felt it was important to search for newly available study data to better represent the simulated population in this REA. In addition, while using the Schofield (1985) equations there were occasional abrupt discontinuities in the estimated RMR observed at some of the equation boundaries (e.g., between age 59 and 60), which were largely a function of how the data were stratified (six age groups and two sexes) and the resulting equation parameters.

Since the 2009 REA, we have reviewed recent RMR literature and other published sources containing individual data and have compiled the associated individual RMR measurements, along with associated influential attributes such as age, sex, and body weight, where available. Data from these individual studies were then combined with RMR data reported in the Oxford-Brookes database (Henry, 2005; IOM, 2005) and screened for duplicate entries. In addition, observations missing values for RMR, BW, age, or sex were deleted, resulting in a dataset containing 16,254 observations (9,377 males and 6,877 females).

Using this new RMR dataset, and having a goal of updating the previous RMR equations and reducing discontinuities in RMR between age groups, new equations were developed. The equations follow the general format of a multiple linear regression (MLR) model, using age and body weight as independent variables to estimate each simulated individual's RMR, along with a residual error term (ε).[20] It is known that RMR and BW, as well as RMR and age, are not exactly linearly related; the algorithms developed here use BW (in kg), age (in years), and the natural logarithms of BW and (age+1)[21] as follows in Equation 4-2, with their parameter estimates provided in Table 4-3.

$$RMR = \beta_0 + \beta_1 \mathrm{BW} + \beta_2 \log(\mathrm{BW}) + \beta_3 Age + \beta_3 \log(Age) + \varepsilon_i \qquad \text{Equation 4-2}$$

When comparing observed versus predicted values, the new RMR equations have a bias of less than 0.5%, compared to the previously used APEX equations which had a bias of between 1-2%. Further, the discontinuities in RMR seen across particular age group boundaries using the previous equations have been reduced when using these updated equations in APEX. Additional

[20] The residual error term largely accounts for the estimation of inter-personal variability in RMR for individuals having the same body weight and age. There are other potentially influential sources of variability that are not explicitly accounted for by the equation (e.g., seasonal influences on RMR) and thus remain as an uncertainty.

[21] The "+1" modifier allows APEX to round age upwards instead of downwards to whole years, which is necessary to avoid undefined log(0) values.

details regarding the derivation of the updated RMR equation and performance evaluation are provided in Appendix H. One standard APEX input file is used for the SO_2 assessment:

- *Physiology040617_noHT_Graham_Glen_QA.txt*: Regression coefficients used to estimate RMR (kcal day^{-1}) for two sexes and six age groups.

Table 4-3. Regression parameters used to estimate RMR by sex and age groups.

Sex	Age Group	n	BW	log(BW)	Age	log(Age)	Intercept	Std dev
male	0–5	625	13.19	270.2	-18.34	131.3	-208.5	69.10
	6–13	1355	10.21	260.2	13.04	-205.7	333.4	115.3
	14–24	4123	0.207	1078.	115.1	-2794.0	3360.6	161.1
	25–54	2531	2.845	729.6	3.181	-191.6	-1067	178.2
	55–99	743	9.291	264.8	-5.288	181.5	-705.9	163.6
female	0–5	625	11.94	261.5	-22.31	120.9	-183.6	64.16
	6–13	1618	5.296	409.1	40.37	-524.9	392.7	99.43
	14–29	2657	0.968	676.9	40.89	-1002	772.7	143.1
	30–53	1346	4.935	355.4	16.28	-896.0	2225	145.3
	54–99	631	2.254	445.9	5.464	-489.9	944.2	124.5
Units: RMR = kilocalories/day; BW = kilograms; Age = years								

Following the estimation of an age- and sex-specific RMR for simulated individuals, the next variable used for estimating ventilation rate involved an approximation of the energy expended for activities an individual performs throughout their day. As mentioned above, activity-specific energy expenditure is highly variable and can be estimated using metabolic equivalents of work (METs), or the ratios of the rate of energy consumption for non-rest activities to the resting metabolic rate of energy consumption, as follows:

$$EE = MET \times RMR \qquad \text{Equation 4-3}$$

where,

EE = Energy expenditure (kcal/minute)

MET = Metabolic equivalent of work (unitless)

RMR = Resting metabolic rate (kcal/minute)

Statistical distributions of METs were developed for simulated activities using the physical-activity compendium (Ainsworth et al., 2011; hereafter “the compendium”). The compendium contains a point value for the MET associated with each of several hundred different activities. Activity-specific MET distributions were developed by cross-walking the activities described in the compendium with the descriptions of activities in the activity pattern data base used by APEX (US EPA, 2017c). The shape of the statistical distribution (e.g., normal, lognormal, triangular, point) for each activity was assigned based on the number of corresponding activities in the compendium and goodness-of-fit statistics. When simulating individuals, APEX randomly samples from the activity-specific METs distributions to obtain values for every activity performed. Two standard APEX input files are used for the SO_2 assessment:

- *MET_Distributions_092915.txt*: MET distribution number, statistical form, distribution parameters, lower and upper bounds, activity description
- *MET_mapping_current_APEX_file.txt:* activity codes, age group (where applicable), occupation group, MET distribution number, and activity description used to link of MET distributions to activities performed

The rate of oxygen consumption ($\dot{V}O_2$, Liters min^{-1}) for each activity is then calculated from the energy expended (kcal min^{-1}) using an energy conversion factor (ECF, Liters O_2 $kcal^{-1}$) as follows in Equation 4-4:

$$\dot{V}O_2 = EE \times ECF \qquad \text{Equation 4-4}$$

The value of the ECF is randomly selected from a uniform distribution for each person, U[0.20, 0.21] (Johnson et al., 2002, adapted from Esmail et al., 1995). One standard APEX input file is used for the SO_2 assessment:

- *Physiology040617_noHT_Graham_Glen_QA.txt*: Parameters of the uniform distribution representing the ECF used for all ages and both sexes.

4.1.3.3 Ventilation Rate

Human activities are variable over time, with a wide range of activities possible within only a single hour of the day. The type of activity an individual performs, such as sleeping or jogging (as well as individual-specific factors such as age, weight, RMR) will influence their ventilation rate. APEX estimates minute-by-minute ventilation rates that account for the expected variability in the activities performed by simulated individuals. Ventilation rate is important in this assessment because the lung function responses associated with short-term peak SO_2 exposures coincide with moderate or greater exertion (ISA, section 5.2.1.2). In our exposure modeling approach, we used APEX to generate the complete time-series of activity-specific ventilation rates and the corresponding time-series of estimated SO_2 exposures. APEX then aggregates both the ventilation rate and exposure concentration to the averaging time of interest (a 5-minute average). Thus, the model provides exposure estimates for the simulated individuals that pertain to specific target levels for both ventilation rate and exposure concentration. The approach to estimating activity-specific energy expenditure and associated ventilation rate involves several algorithms and physiological variables, with details found in the APEX User's Guide (U.S. EPA, 2017a, b).

Using the existing measurement $\dot{V}_E$ dataset from Graham and McCurdy (2005), new $\dot{V}_E$ algorithms were developed for predicting activity specific $\dot{V}_E$ in the individuals simulated by APEX. The new $\dot{V}_E$ algorithms do not directly employ previously used variables to stratify the data (age groups, sex) and explain variability (age, body weight, height) in ventilation rate, effectively simplifying and reducing the number of equations. The new algorithms utilize a new variable, the maximum volume of oxygen consumed ($\dot{V}O_2m$) as an input.[22] Body weight, height and sex – as well as fitness level (which is often represented by $\dot{V}O_2m$) - influence oxygen consumption for a particular activity. However, variability for each of these influential variables are already captured in the algorithm used to estimate each simulated individual's RMR, and subsequently, the estimation of their activity specific $\dot{V}O_2$.[23] Thus, the only input variables needed for the new $\dot{V}_E$ algorithm are $\dot{V}O_2$ and $\dot{V}O_2m$,[24] both of which are estimated by APEX.

Details for the derivation of and performance evaluation of the new equation that APEX uses to estimate ventilation rate are provided in Appendix H. Briefly, the $\dot{V}_E$ dataset contains 6,636 observations, with 4,565 males and 2,071 females. Similar to the earlier ventilation equation by Graham and McCurdy (2005), a mixed-effects regression (MER) model was fit because the MER separates residuals into within-person (e_w) and between-person (e_b) effects, known as intrapersonal and interpersonal effects, respectively.[25] It was found that the actual values of $\dot{V}O_2$ and $\dot{V}O_2m$ are less relevant than the fraction of maximum capacity, represented by $f_1 = \dot{V}O_2/\ \dot{V}O_2m$. The variable f_1 may operate non-linearly (for example, $f_1 = 0.9$ is likely *more* than twice as encumbering as $f_1 = 0.45$). A transformation regression approach (PROC TRANSREG – SAS, 2017) was used to determine the most appropriate variable transformation, indicating a power of 4 to 5 be used when only the log transformed $\dot{V}O_2$ was used as the independent variable and described in Equation 4-5.

[22] Use of $\dot{V}O_2m$ as an explanatory variable in separate related research on metabolic equivalents of task (MET) values for persons with unusual maximum capacity for work suggests that their MET distributions are modified in a predictable way by their maximum MET (or, equivalently, by $\dot{V}O_2m$), thus providing support for use of this variable in the new $\dot{V}E$ algorithms.

[23] Oxygen consumption associated with activities performed is based on the activity specific metabolic equivalents for work (METs), an individual's estimated RMR, and an energy to oxygen conversion factor (U.S. EPA, 2017b).

[24] Distributions of VO_2m used by APEX were derived from 20 published studies reporting individual data and grouped mean (and standard deviation) data obtained from 136 published studies. Details are provided in Isaacs and Smith (2005).

[25] $N(0, e_b)$ is a normal distribution with mean zero and standard deviation e_b=0.09866 meant to capture *inter*personal variability, which is sampled once per person. $N(0, e_w)$ is an *intra*personal residual with standard deviation of e_w=0.07852, which is resampled daily due to natural *intra*personal fluctuations in $\dot{V}_E$ that occur daily.

$$\dot{V}_E = e^{(3.300 + 0.8128 \times \ln(\dot{V}O_2) + 0.5126 \times (\dot{V}O_2 \div \dot{V}O_2 m)^4 + N(0,eb) + N(0,ew))} \qquad \text{Equation 4-5}$$

In comparing the statistical fit of the new equation with the equations used by APEX previously to estimate ventilation rate, the resulting coefficient of determination (r^2 values) for the new equation ($r^2 = 0.94$) indicates an improved fit compared to that of the previous equations ($r^2 = 0.89$-0.92). Further, because the data were not stratified by age groups (or any other groupings), there are no discontinuities in predictions made across age boundaries as was observed when employing the previous equation. Information used in estimating ventilation rate is found in the following APEX two input files:

- *Physiology040617_noHT_Graham_Glen_QA.txt*: parameters describing statistical distributions of normalized maximum oxygen consumption rate ($N\dot{V}O_2m$) for two sexes by single age years (0-100) (see, Isaacs et al., 2005).
- *Ventilation_VEMethod2_102816_new.txt*: minimum and maximum age ranges, regression coefficients, between and within error terms used to estimate individual activity-specific ventilation.

To use this information to estimate health risks for children, the ventilation rates observed for the adult study subjects need to be converted into rates that best reflect the different physiology of children. Consistent with prior REAs (U.S. EPA, 2009, 2014b; Whitfield et al., 1996) we used an equivalent ventilation rate (EVR), which is essentially an allometrically normalized ventilation rate, to estimate instances when a simulated individual reaches a ventilation rate relatively as high as that of the study subjects (i.e., termed here as moderate or greater exertion).

$$EVR = \frac{\dot{V}_E}{BSA} \qquad \text{Equation 4-6}$$

In the controlled human exposure studies evaluating the health effects of SO_2, the ventilation rates for study subjects (i.e., male and female adults) experiencing effects from 5- to 10-minute SO_2 exposures are generally within 40-50 L/min, with most set at or around 40 L/min (ISA, Table 5-2 and Table 4-12 below).[26] However, body surface area was not measured in the controlled human exposure studies and the relevant ventilation data were not separated by sex. We approximated BSA of the study subjects as 1.82 m^2 based on data for adult males and

[26] In these studies, subjects were breathing freely during exercise; thus, it is expected that there was a mixture of nasal, oral, and oro-nasal breathing that occurred across the study subjects. Without information regarding the precise breathing method used by any subject corresponding with their health response, we assumed that the mixture in breathing method used by study subjects is representative for the simulated population.

females from U.S. EPA (1989).[27] Based on these data, we estimate EVR for the study subjects to be 40/1.82 ≈ 22 L/min-m^2. Accordingly, we have used this EVR as the target EVR in this assessment and simulated individuals at or above an EVR of 22 L/min-m^2 (children or adult) during a 5-minute exposure event were characterized as performing activities at or above moderate exertion. This is essentially the same target EVR value as that used in the 2009 REA (i.e., ≥ 22 L/min-m^2), approximated at that time based on data from U.S. EPA (1997). This value used for EVR would represent a mean value based on the data used in its estimation and is considered reasonable to apply and approximate when, on average, individuals in a population might experience periods of moderate or greater exertion. There is uncertainty in this mean based on the broad scale of data used in it derivation, as well as by not having any information on how to characterize intra- or inter-personal variability, where existing, and in its direct extrapolation from adults to children.

4.2 METEOROLOGICAL DATA

Temperature data are used by APEX in selecting human activity data and in estimating AERs for indoor residential MEs. Hourly surface temperature measurements were obtained from the National Weather Service Integrated Surface Hourly (ISH) data files (described in section 3.2.1.1). The weather stations used for each study area are given in Table 4-4. Given the limited geographic area of each study area, data from a single station was used to represent the ambient air temperature in each study area. The occurrence of missing temperature data was limited to a few instances (Table 4-4). Temperature values for the hours missing data were estimated using SAS PROC EXPAND, a simple linear interpolation technique. Because of the small number of missing values, the impact of the filled values to estimated exposures is assumed negligible. Multiple unique APEX input files are used, one for each year and study area, and generally in two formats:

- *METdata[studyarea]Y[year].txt*: MET station IDs, hour of day, hourly temperature (°F) for each MET station, by study area and year
- *METlocs[studyarea]Y[year].txt:* MET station ID's, latitudes and longitudes, start and stop dates of temperature data

[27] Most of the controlled human exposure studies were conducted in the 1980s, thus use of the 1989 EPA Exposure Factors Handbook is considered the most representative source to use in estimating BSA for the study subjects compared with the 1997 and 2011 versions of that document given that body weight distributions (and hence BSA) have changed over time.

Table 4-4. Study area meteorological stations, locations, and hours of missing data.

Study Area	Station Name	Station Number	Latitude	Longitude	Number of hours with missing temperature		
					2011	2012	2013
Fall River, MA	PROVIDENCE T F GREEN ARPT	14765	41.7225	-71.4325	0	0	5
Indianapolis, IN	INDIANAPOLIS INTERNATIONAL APT	93819	39.72517	-86.28168	0	0	0
Tulsa, OK	RICHARD LLOYD JONES JR APT	53908	36.0396	-95.9846	10	0	0

4.3 CONSTRUCTION OF HUMAN ACTIVITY SEQUENCES

Exposure models use human activity pattern data to predict and estimate exposure to pollutants. Different human activities, such as outdoor exercise, indoor reading, or driving a motor vehicle can lead to different pollutant exposures. This may result from differences in the amount of the pollutant in the different locations where the activities are performed as well as from differences in the energy expended in performing the different activities (because energy expenditure influences inhalation and ingestion and thus may influence pollutant intake). To accurately model exposures to ambient air pollutants, it is critical to have a firm understanding of the locations where people spend time and the activities performed in such locations. The following subsections describe the activity pattern data, population commuting data, and the approaches used to simulate where individuals might be and what they might be doing.

After the basic demographic variables are identified by APEX for a simulated individual in the study area, values for the other variables are selected as well as the development of the activity patterns that account for the places the simulated individual visits and the activities they perform. The following subsections describe the population data we used in the assessment to assign key features of the simulated individuals, and approaches used to simulate the basic physiological functions important to the exposure estimates for this REA.

4.3.1 Consolidated Human Activity Database

The Consolidated Human Activity Database (CHAD) provides time series data on human activities through a database system of collected human diaries, or daily time location activity logs (U.S. EPA, 2017c). The purpose of CHAD is to provide a basis for conducting multi-route, multi-media exposure assessments (McCurdy et al., 2000). The data contained within CHAD come from multiple surveys with variable study-specific structure (e.g., real time minute-by-minute recording of diary events versus a recall method using time-block-averaging). Common to all studies, individuals provided information on their locations visited and activities performed for each survey day. Personal attribute data for these surveyed individuals, such as age and gender, are included in CHAD as well. The latest version of CHAD contains data for nearly 180,000 person-days, however for this assessment, APEX uses about 55,000 of these.[28] Most of the CHAD data are from studies conducted since 2000, several of which are newly included since the 2009 REA. See Appendix I for a list of the studies available, study dates, and number of diaries included from each. Three standard APEX input files are used for the current assessment to create the activity pattern profiles for all simulated individuals.

- *Activity_diaries_events_no_ATUS_BLS.txt:* CHAD ID, clock hour (hhmm), duration of event (minutes), CHAD activity code, and CHAD location code, serving as a daily sequence of locations visited, activities performed, and their duration for individuals in CHAD
- *Activity_diaries_questionnaire_no_ATUS_BLS.txt*: CHAD ID, day-of-week, sex, race, employment status, age, maximum daily temperature, average temperature, occupation, missing time (minutes), record count, commute time (see also section 4.3.2)
- *Activity_diaries_statistics_no_ATUS_BLS.txt*: CHAD ID, total daily time spent outdoors (minutes) (see also section 4.3.4)

4.3.2 Commuting Activity Pattern Data

Exposures can vary across a study area based on spatial heterogeneity in ambient air concentrations and how that corresponds with a simulated individual's activity pattern and geographic location. APEX approximates home-to-work commuting flows between census designated areas for each employed individual, and thus accounts for differing ambient air concentrations that may occur in these geographic locations. APEX has a national commuting database originally derived from 2010 Census tract level data collected as part of the U.S. DOT Census Transportation Planning Package. The data used to generate the APEX commuting file are from the "Part 3-The Journey to Work" files.[29] The Census files contain counts of individuals commuting from home to work locations at a number of geographic scales. These data have been

---

[28] Data from the U.S. Bureau of Labor Statistics American Time Use Survey (ATUS) are in CHAD master 071113, but they are not used by APEX in our simulations because of an important survey coding issue. Time spent at home for ATUS participants was not distinguished as indoors or outdoors, an important distinction for accurately estimating SO_2 exposures. It could be possible to approximate the time expenditure of the ATUS diaries using an independent source of information, such as using the other CHAD diaries that recorded indoor and outdoor time (e.g., the 55,000 CHAD diaries used for estimating exposure would be the best source of information). However, it is unlikely that the representation of time expenditure would change/improve nor would the estimated exposures differ when including modified ATUS diaries that would reflect the same pattern in indoor and outdoor time as the 55,000 CHAD diaries already used in our exposure simulations.

[29] These data are available from the U.S. DOT Bureau of Transportation Statistics http://transtats.bts.gov/) at the web site: https://www.transtats.bts.gov/Fields.asp.

processed to calculate fractions (and hence commute probabilities) for each tract-to-tract flow to create the national commuting data distributed with APEX. This database contains commuting data for each of the 50 states and Washington, D.C. This data set does not differentiate people that work at home from those that commute within their home tract. A companion file to the commuting flow file is the commuting times file, i.e., an estimate of the usual amount of time in minutes it takes for commuters to get from home to work each day and tract-to-tract commuting distances.[30] The commuting times file information is used to select CHAD activity pattern data from individuals having time spent inside vehicles similar to the census commute times and associated distances travelled. To use these tract level files at the block level for this REA, all blocks were assumed by APEX to have the same commuting probabilities as the parent tract for commuting to the blocks within other tracts by using the 11-character identifier common to both IDs. Intra-block (within a tract) commuting is unknown and thus not simulated. Two standard APEX input files are used for the current assessment, as listed here.

- *CommutingTimesBlock2010_3StudyAreas.txt*: census block ID's, count of all employed individuals, count of employed individuals that do not work at home, 7 groups of block-level one-way commuting times (in minutes)[31]
- *Commuting_flow_US_2010_tracts.txt*: census tract IDs, tract-to-tract commute cumulative probabilities (in fractional form), commute distance (km)

4.3.3 Assigning Activity Pattern Data to Individuals

Once APEX identifies the basic personal attributes of a simulated individual (section 4.1) and daily temperatures (section 4.2), activity pattern data from CHAD are selected based on age,[32] sex, temperature category, and day of the week. These attributes are considered first-order attributes in selecting CHAD diaries when modeling human exposures (Graham and McCurdy, 2004). The maximum daily temperature range is used to select activity pattern data that best match the study area meteorological data for the simulated individual. This information is found in the following APEX input file, varying by study area and simulation year:

- *Functions_ [studyarea]Y[year].txt:* probabilities and interval definitions associated with a few input variables. For activity diary selection - day of week intervals (weekend or weekday) by three temperature ranges (<55, 55-83, or >83 °F).

[30] These data are from the U.S. Census data portal (http://dataferrett.census.gov/) and are found in Table P31, variables P031001-P031015.

[31] The nine commuting time groups in this file are: 0-4, 5-14, 15-19, 20-29, 30-44, 45-59, and >60 minutes.

[32] Rather than select using exact ages, APEX allows the user to expand the pool of available diaries using the variable 'AgeCutpct' which allows for diaries to meet the simulated individuals required age within a certain percent of that age. A value of 15% was selected (with a default minimum of 1 year). For instance, CHAD diaries from people ages 51 to 69 would be available to simulate a person aged 60.

While there may be other important attributes that may influence activity patterns (e.g., obesity, disease status), there are limits to our ability to link to all the possible personal attributes that may be of interest in modeling an individual's activities to the CHAD data. This is largely because CHAD is a compilation of data collected from numerous individual activity pattern studies conducted over several decades, many of which had a unique survey design. As a result, there is a varying amount of missing personal attribute data for the CHAD diaries. For instance, there are only a limited number of CHAD diaries with survey-requested health information (e.g., the health status of respondents). Specifically regarding whether or not a survey participant had asthma, about 70% of the available diaries used by APEX in this REA had either a 'yes' or 'no' response to this health condition, of which there were 5,107 diary days representing individuals having asthma (of which 3,734 were children). This may appear to be a large number of diaries, however, following a grouping of the diaries by their first-order attributes (i.e., stratifying and reducing the available data for these diary groups of 'like individuals' by about a factor of 20 or so), there would be fewer than 200 diaries available for simulating a single day for that particular individual. Accordingly, the selection of diaries to use for APEX-simulated individuals does not consider health status (e.g., whether they were for people specifying they did or did not have asthma, or whether such information was indicated by the survey participant).

This restriction in the number of diaries from individuals having asthma is not considered to be a significant limitation for estimating exposures for simulated individuals with asthma in this REA. In general, modeling people with asthma similarly to healthy individuals (i.e., using the same time-location-activity profiles) is supported by the activity analyses reported by van Gent et al. (2007) and Santuz et al. (1997). Other researchers, for example, Ford et al. (2003), have shown significantly lower leisure time activity levels in asthmatics when compared with individuals who have never had asthma. Based on these conflicting conclusion, we evaluated this issue in the 2014 O_3 REA[33] and, using the available activity pattern data in the CHAD database, we compared participation in afternoon outdoor activities at elevated exertion levels among people having asthma, people not having asthma, and unknown health status. The 2014 O_3 REA analysis and associated conclusions are described below.

As is of interest in this current SO_2 REA, we wanted to focus on instances when individuals would experience their highest O_3 exposures. As has been shown in SO_2 and O_3 exposure assessments (U.S. EPA, 2009; U.S. EPA, 2014), the highest exposures occur when

[33] See 2014 O_3 REA sections 5.4.1.5 and 5G-1.4 for details (U.S. EPA, 2014). While there are about 8,300 more diaries in the CHAD used for this SO_2 REA, about 5,800 of the additional diaries added since the 2014 O_3 REA have an unknown health status. Note, the percent of diaries from people with asthma is nearly identical in both data sets: children with asthma - 20.6% in 2014 O_3 CHAD vs 20.3% in 2018 SO_2 CHAD; adults with asthma – 7.5% in 2014 O_3 CHAD vs 7.5% in 2018 SO_2 CHAD. Therefore, rather than generate a new evaluation in this REA, conclusions drawn from the prior analysis are considered reasonable for this REA.

individuals spend time outdoors, particularly during the afternoon hours. To prepare the data set for analysis, afternoon hours were characterized as the time between 12 PM and 8 PM and only those persons that spent some time outdoors were retained. As is done by APEX in simulating individuals, level of exertion was estimated by sampling from the specific METS distributions assigned for each person's activity performed. Then, we identified activities having a METS value of greater than 3 as times where a person was at moderate or greater exertion levels (US DHHS, 1999). Afternoon outdoor time was then stratified by exertion level, summed for two study groups of interest (children and adults), and presented in percent form within Table 4-5.

Of the CHAD diaries for children, about 18% are from an individual with asthma and 69% are from those who do not have asthma. About 5% of CHAD diaries for adults are from individuals with asthma, and about 65% are from those who do not have asthma. Far fewer children's diaries are from persons whose asthma status is unknown (12%) compared to adults (30%), and the proportions are smaller still in terms of the total available person-days. On average, about 43% of all children of known asthma status spent some afternoon time outdoors, and the percent is actually higher for children with asthma (48.5%) than for children not having asthma (41.2%). About half of the adults whose asthma status was known spent afternoon time outdoors with a participation rate generally similar for adults having asthma and adults not having asthma. Participation in outdoor events for persons having unknown asthma status varied from that of persons with known asthma status; about 60% of the children's diaries with unknown asthma status and 31% of the adult diaries indicate some afternoon time was spent outdoors.

Table 4-5. Comparison of outdoor time expenditure and exertion level by asthma status for children and adult CHAD diaries used by APEX.

	CHAD: Children (4 to 18) [a]			CHAD: Adults (19 to 95) [b]		
Has Asthma?	Yes	No	Unknown	Yes	No	Unknown
Total Person Days (n)	3,206	12,346	2,128	1,254	15,465	7,075
Number of Person Days with Time Spent Outdoors (% participation)	1,564 (48.8%)	5,092 (41.2%)	1,267 (59.5%)	602 (48.0%)	7,949 (51.4%)	2,176 (30.8%)
Percent of Afternoon Hours Spent Outdoors (%)	28.5%	27.5%	28.9%	26.2%	27.2%	22.2%
Percent of Afternoon Time Outdoors at Moderate or Greater Exertion (%)	80.3%	78.2%	79.2%	62.7%	63.8%	60.3%

From Table 5G-2 of 2014 O_3 REA (U.S. EPA, 2014)
[a] CHAD studies for where a survey questionnaire response of whether or not child was asthmatic include CIN, ISR, NHA, NHW, OAB, and SEA (see Appendix I for study names).
[b] CHAD studies for where survey a questionnaire response of whether or not adult was asthmatic include CIN, EPA, ISR, NHA, NHW, NSA, and SEA.

The amount of time spent outdoors by the persons that did so varied little across the two study groups and three asthma classifications. On average, diaries from children indicate approximately 2¼ hours of afternoon time is spent outdoors, 80% of which is at a moderate or greater exertion level, regardless of their asthma status. For individuals whose asthma status is known, slightly less afternoon time is spent outdoors by adults (about 125-130 minutes) than children and the percent of afternoon time adults perform moderate or greater exertion level activities is also lower (about 63%). As noted above regarding the reduced participation in outdoor events for adults whose asthma status is unknown, diaries for these adults also have about 20 fewer minutes of afternoon time spent outdoors compared with those persons whose asthma status is known. Based on this analysis and additional comparisons of CHAD diary days with literature reported values of outdoor time participation at varying activity levels (see U.S. EPA, 2014), the 2014 O_3 REA evaluation of the CHAD data indicates there are strong similarities in outdoor time, outdoor event participation, and activity levels among the three study groups and with those reported in independent studies of people with asthma. Thus, we conclude the use of any CHAD diary, regardless of asthma status, is reasonable for purposes of simulating people with asthma in this exposure assessment.

4.3.4 Method for Longitudinal Activity Sequences

In order to estimate population exposure over a full year, a year-long activity sequence needed to be created for each simulated individual based on CHAD, which is largely a cross-sectional activity database of 24-hour records. The typical surveyed subject in the time location activity studies in CHAD provided about two days of diary data. For this reason, the construction of a season-long activity sequence for each individual requires some combination of repeating the same data from one subject and using data from multiple subjects. The best approach would reasonably account for the day-to-day and week-to-week repetition of activities common to individuals, and recognizing even these diary sequences are not entirely correlated, while maintaining realistic variability among individuals comprising each study group.

APEX provides three methods of assembling composite diaries. We have selected the method for this assessment based on our consideration of the assessment objectives, consideration of an evaluation of differences in results produced by the three methods and consideration of flexibility provided by each approach with regard to specifying key variable values. Based on all of these considerations, we have selected the D&A method.

The D&A method is a complex algorithm for assembling longitudinal diaries that attempts to realistically simulate day-to-day (within-person correlations) and between-person variation in activity patterns (and thus exposures). This method was designed to capture the tendency of individuals to repeat activities, based on reproducing realistic variation in a key diary variable, which is a user selected function of diary variables. The method targets two statistics: a population diversity statistic (D) and a within-person autocorrelation statistic (A). The D statistic reflects the relative importance of within and between-person variance in the key variable. The A statistic quantifies the lag-one (day-to-day) key variable autocorrelation. Values of D and A for the key variable are selected by the model user and set in the APEX parameters file, and the method algorithm constructs longitudinal diaries that preserve these parameters. Further details regarding this methodology can be found in Glen et al. (2008).

Besides the D&A method, there are two additional methods of compiling diaries provided by APEX: a more basic method and a similarly complex method. The more basic method involves randomly selecting an appropriate activity diary for the simulated individual from the available diary pool. While this more basic method is adequate for providing a mean short-term exposure estimate, it is less useful for this assessment for which the objective is to estimate how often individuals may experience particular peak SO_2 exposures over a year. The more complex method uses a Markov-chain clustering (MCC) approach in attempting to recreate realistic patterns of day-to-day variability. First, cluster analysis is employed to divide the daily activity pattern records into three groups based on time spent in five microenvironments: indoor-residence, other indoors, outdoor-near roads, other outdoors, and inside vehicles. For each simulated individual, a single time-activity record is randomly selected from each cluster. Then the Markov process determines the probability of a given time-activity pattern occurring on a given day based on the time-activity pattern of the previous day and cluster-to-cluster transition probabilities (and are estimated from the available multi-day time-activity records), thus constructing a long-term sequence for a simulated individual. Details regarding the MCC method and supporting evaluations are provided in the 2009 REA Appendix B, Attachments 4 and 5.

Che et al. (2014) performed an evaluation of the impact of the three APEX methods on $PM_{2.5}$ exposure estimates. As expected, little difference was observed across the methods with regard to estimates of the mean exposures of simulated individuals. Differences were observed, however, in the number of multiday exposures exceeding a selected benchmark concentration. With regard to the number of simulated individuals experiencing 3 or more days above benchmark concentrations, the MCC method estimates were approximately 12-14% greater than either the random or D&A methods. For the number of persons experiencing at least one exposure of concern, however, the MCC method estimates were approximately 4% lower than those of the other two methods. For additional context, we note that, using all methods, there is an order of magnitude difference in the number of persons exposed at least once versus three or more times, indicating that, overall, the occurrence of simulated multiday exposures are rare events regardless of method selection.

Che et al. (2014) concludes that while the MCC method produces a higher number of multiday exposures, there remains a question whether the MCC method has greater accuracy relative to the other two methods. We note this conclusion applies to both the estimations of single day and multiday exposures, as there is an inverse relationship between the two when simulating exposures using APEX and a finite set of activity pattern data. Thus, the MCC method produces a smaller number of single day exposures above benchmarks relative to the other two methods, estimations also subject to a degree of uncertainty.

In the absence of having a robust data set (e.g., multiday/week personal exposure information from a random population) to better evaluate the accuracy of any of the methods, we considered selection of the longitudinal approach for this assessment from a practical perspective, guided by a balancing of the single day and multiday exposures that can be estimated by each method. In so doing, we selected the D&A approach, recognizing that the D&A method allows for flexibility in the selection of the key influential variable and its setting values, and also the ability to directly observe the impact of changes to these values on model outputs.

The key variable selected for this REA is the amount of time an individual spends outdoors each day, as that is one of the most important determinants of exposure to high levels of SO_2 (see section 2.1.2 above). In their evaluation, Che et al. (2014) varied the values of D and A for this variable to determine the impact to estimated exposures. Compared to the base level simulation (i.e., D=0.19 and A=0.22),[34] increasing both D and A by 100% increased the number of persons having at least three exposures above the selected benchmark by about 4%, while also reducing the percent of persons experiencing at least one day above benchmarks by less than 1% (Che et al., 2014). In recognizing uncertainty in the parameterization of D and A (i.e., based on a limited field study of a small subset of the population, children 7-12) and that the base level simulation D&A values produced a lower estimate of repeated exposures compared with the MCC method, we have used values of 0.38 for D and 0.44 for A for all ages to potentially increase representation of multiday exposures without significant reducing the percent of the population experiencing at least one day at or above benchmark concentrations.

4.4 MICROENVIRONMENTAL CONCENTRATIONS

In APEX, exposure of simulated individuals occurs in microenvironments. To best estimate personal exposures, it is important to maintain the spatial and temporal sequence of microenvironments people inhabit and appropriately represent the time series of concentrations

[34] Longitudinal diary data from a limited field study of children ages 7-12 (Geyh et al. 2000; Xue et al. 2004) provide support for estimates of approximately 0.19 for D and 0.22 for A for the amount of time spent outdoors.

that occur within them. Two methods available in APEX for calculating pollutant concentrations within microenvironments are a mass balance model and a transfer factor approach. In both approaches, ME concentrations depend on the ambient (outdoor) air SO_2 concentrations and temperatures, as well as distributions of the key parameters for each approach. Further, the distributions of some of the key parameters depend on values of other variables in the model. For example, the distribution of air exchange rates inside an individual's residence depends on the type of heating and air conditioning present, which are also stochastic inputs to the model. The value of a stochastic parameter can be set as a constant for the entire simulation (e.g., house volume would remain identical throughout the exposure period), or APEX can be directed to sample a new value hourly, daily, or seasonally from specified distributions. APEX also allows the user to specify diurnal, weekly, or seasonal patterns for certain ME parameters.

Based on findings from the 2009 REA, we have specified five MEs for use in this assessment, largely based on two factors: the expectation of an ME having exposure concentrations of interest and the availability of factors to reasonably model the ME. The 2009 REA results indicated that the majority (70-90%) of 5-minute daily maximum SO_2 exposures between 100 and 800 ppb[35] occurred while individuals were within outdoor microenvironments (2009 REA, Figure 8-21). Given that finding and the objective for this assessment (i.e., understanding how often and where short-term peak SO_2 exposures occur), we recognized the added efficiency of minimizing the number of MEs, particularly lower-exposure indoor MEs, that were parameterized and included in the modeling.

Accordingly, we aggregated the number of MEs to address exposures of ambient air origin that occur within a core group of indoor, outdoor, and vehicle MEs. It was expected that the exposures occurring near roads would also be associated with high exposures as they would be modeled identically to all of the other outdoor MEs, only that these outdoor events occur near a road. Thus, the near road ME was modeled separately in case time spent in that ME and its associated exposures was of specific interest. An inside-vehicle ME was also modeled based on the expectation that it would lead to some instances of high exposures, particularly considering the high air exchange rate that occurs inside vehicles while moving and having a limited SO_2 decay rate, effectively reflecting similar concentration levels as in outdoor MEs. Two indoor MEs (indoor-residence and indoor-other) were modeled individually based on having specific air exchange rate data available for each (4.4.1 and 4.4.3, respectively). The indoor-other ME is comprised of all non-residential MEs, thus could include workplaces or office buildings, stores

[35] Although these results were associated with a different air quality scenario than is evaluated in this REA, the similarity in the scenario leads us to conclude the results are relevant for judgments made here. Air quality in the 2009 REA results referenced here was adjusted to just meet a 99th percentile 1-hour daily maximum single year standard level of 150 ppb (U.S. EPA, 2009).

for shopping, medical offices, and so on. Table 4-6 lists the five microenvironments selected for this analysis and the exposure calculation method for each. The variables used and their associated parameters to calculate ME concentrations are summarized in subsequent subsections below. Details on the calculation of ME concentrations in APEX are presented in Appendix F, section F.7.

Table 4-6. Microenvironments modeled and calculation method used.

Microenvironment (ME)	APEX ME Number	Calculation Method	Variables [a]
Indoor – Residence	1	Mass balance	AER & RM
Indoor – Other	2	Mass balance	AER & RM
Outdoor	3	Factors	None
Near-road	4	Factors	None
Vehicle	5	Factors	PE

[a] AER = air exchange rate, RM = removal rate, PE = fraction of ambient pollutant entering microenvironment,
None = ME concentration is equal to ambient concentration

The mass balance method, used for the indoor MEs, assumes that an enclosed microenvironment (e.g., a room within a home) is a single well-mixed volume in which the air concentration is approximately spatially uniform (Figure 4-3). The concentration of an air pollutant in such a microenvironment is estimated using (1) inflow of air into the microenvironment, (2) outflow of air from the microenvironment, (3) removal of a pollutant from the microenvironment due to deposition, filtration, and chemical degradation, and (4) emissions from sources of a pollutant inside the microenvironment (not used for this REA).

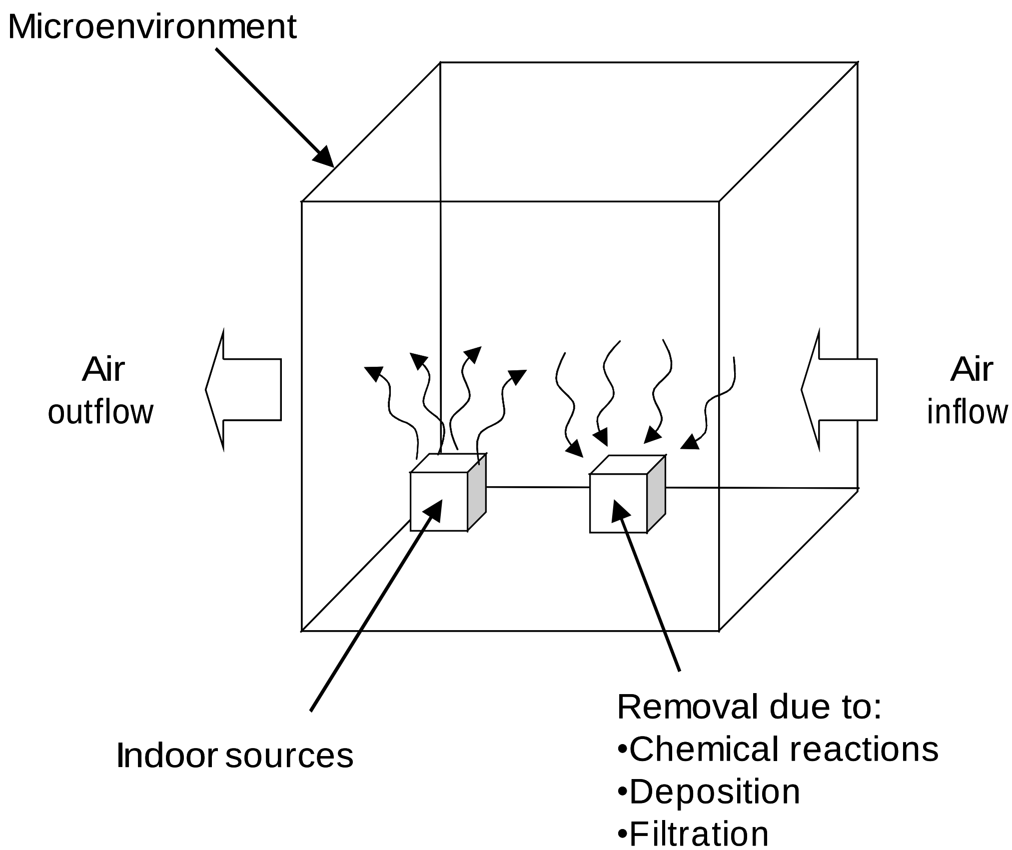


Figure 4-3. Illustration of the mass balance model used by APEX.

Considering the microenvironment as a well-mixed fixed volume of air, the mass balance equation for a pollutant in the microenvironment can be written in terms of concentration:

$$\frac{dC(t)}{dt} = \dot{C}_{in} - \dot{C}_{out} - \dot{C}_{removal} \qquad \text{Equation 4-7}$$

where,

$C(t)$ = Concentration in the microenvironment at time t

$\dot{C}_{in}$ = Rate of change in $C(t)$ due to air entering the microenvironment

$\dot{C}_{out}$ = Rate of change in $C(t)$ due to air leaving the microenvironment

$\dot{C}_{removal}$ = Rate of change in $C(t)$ due to all internal removal processes

The method used for the outdoor MEs uses a factors model and is simpler than the mass balance model. In this method, the value of the concentration in a microenvironment is not dependent on the concentration during the previous time step. Rather, this model uses the Equation 4-8 to calculate the concentration in a microenvironment from the user-provided hourly air quality data:

$$C_{mean} = C_{ambient} \times f_{proximity} \times f_{pollutant} \qquad \text{Equation 4-8}$$

where,

C_{mean} = Mean concentration over the time step in a microenvironment (ppb)

$C_{ambient}$ = The concentration in the ambient (outdoor) environment (ppb)

$f_{proximity}$ = Proximity factor (unitless)

$f_{pollutant}$ = fraction of ambient pollutant entering microenvironment (unitless)

The five microenvironments were mapped to the 115 CHAD location codes,[36] many of which go beyond the scale of the microenvironmental modeling. The ambient air concentration used in calculating ME concentration for each event varies temporally and spatially. For example, commuters (i.e., employed individuals who do not work at home) are assigned to either their home grid or work grid concentrations, depending on whether the population probabilities and commuting data base produce either a home or work event. Additionally, depending on the particular microenvironment (i.e., other than home or work), the mapping of CHAD locations to the five microenvironments also uses an identifier that designates the relative location in the air quality surface from which the ambient air concentration (used to calculate the ME concentration) is selected. For this assessment, such locations would include the blocks from a simulated individual's home (H), work (W), near work (NW), near home (NH), last (L, either NH or NW), other (O, average of all), or unknown (U, last ME determined) census block location. Specific designations are provided in the ME mapping file, with selection based on known factors and professional judgement. For example, when an individual is in their home, the ambient concentration in the home block is used to calculate their ME concentration. When the individual is at work, the block the individual commuted to is used to calculate their ME concentration. Travel inside vehicles used the ambient concentration data from the block used to calculate the prior ME concentration. Most other MEs (both indoor and outdoor) used ambient concentration data selected from near home blocks.

Status attribute variables are also important in estimating ME concentrations, and can include, but are not limited to, housing type, whether the house has air conditioning, and whether the car has air conditioning. Because outdoor MEs are expected to contribute the most to an individuals' highest SO_2 exposure (and potential health risk) and the status attribute variables pertain to indoor MEs, the setting of these particular variables will have limited impact to this REA's exposure results. In this assessment, a number of temperature ranges are used in selecting

[36] The location codes indicate specific MEs that extend beyond simple aggregations of indoor, in-vehicle, and outdoor locations where people spend time. For example, CHAD has a location code for when individuals spent time inside their residence while in the kitchen.

the particular distribution for estimating air exchange rates (AERs). Maximum daily temperature is also used in diary selection to best match the study area meteorological data for the simulated individual (Graham and McCurdy, 2004), and air conditioning use prevalence data.

Multiple APEX ME input files (the first and third in the list below), of the same general format, are used for each study area in the REA. A single ME mapping file is used for all study areas. These files contain the specific setting of all variables described in this section.

- *ME_descriptions_[studyarea]_5MEs.txt*: defines ME calculation method, conditional variables used (e.g., temperature categories – see functions file), distribution type, distribution parameters (mean, standard deviation, minimum, maximum) used for estimating AER, decay rates, proximity factors, and PE fraction to estimate microenvironmental concentrations.
- *MicroEnv_Mapping_CHAD_to_APEX_5MEs.txt*: maps 115 CHAD locations to 5 APEX simulated microenvironments and assigns block-level ambient concentrations to use for each location. Contains CHAD location code, CHAD description, APEX ME number, and ambient concentration location identifier
- *Functions_ [studyarea]Y[year].txt:* variables used for selecting air exchange rates (AER) - air conditioning (A/C) prevalence (home has A/C, does not have A/C) by five temperature ranges for air exchange rate (<50, 50-67, 68-76, 77-85, or >85 °F). (see section 4.4.1 and 4.4.2)

4.4.1 Air Exchange Rates for Indoor Residential Microenvironments

Distributions of AERs for the indoor residential MEs were developed previously using data from several studies. The analysis of these data and the development of most of the distributions used in the modeling, originally described in detail in U.S. EPA (2007) Appendix A and recently updated by Cohen et al. (2012), are provided in U.S. EPA (2014) Appendix 5E.

Briefly, these prior analyses indicated that the AER distributions for the residential MEs depend on the presence or absence of mechanical air conditioning (A/C) and the outdoor temperature (and a few other variables[37] for which sufficient data are not available). Further, the AER distributions vary across the U.S. study locations, [38] such that the selected AER distributions for the modeled study areas should also depend on these influential factors. For each combination of air conditioner (A/C) prevalence, U.S. geographic region (and hence climate zone), and temperature (where data were available), lognormal distributions were fit.

There were a number of limitations in generating study-area-specific AERs, stratified by temperature range and A/C type. For example, the AER data collected and the distributions

---

[37] For example, there were insufficient data available across the studies to indicate the specific A/C unit type (central, window, or both), whether windows were closed or open, or whether a mechanical fan was in operation.

[38] The studies were conducted in several U.S. cities (e.g., Detroit, Houston, Los Angeles, New York), likely accounting for AER differences due to local climate and variability in overall housing stock (e.g., types of residences, year built).

subsequently derived from them were available only for selected cities that had limited numbers of samples collected at varying ambient air temperatures, and yet the summary statistics and comparisons demonstrate that the AER distributions depend upon the city as well as the temperature range and A/C type. Because specific AER data are not available for the study areas in this assessment, we used AER data from Cohen et al. (2012) for a city within the same geographic region as the particular study area, and considering the same temperature ranges on which the AER distributions were originally based. The AER distributions used for the exposure modeling are given in Table 4-7 (for residences with A/C) and Table 4-8 (for residences without A/C). Upper and lower bounds were selected to guard against the generation of extreme AER values. In general, the AER distributions are highest for the Fall River study area, while the AER distributions used for the Tulsa study area are lowest. This implies indoor residential exposures would tend to be greatest for the simulated individuals in the Fall River study area when compared with the other two study areas, though again, the expectation is that outdoor exposures would contribute to the highest exposures in all three study areas and limit the importance of these observed differences in AER.

Table 4-7. AERs for indoor residential microenvironments (ME-1) with A/C by study area and temperature.

Study Area	Daily Mean Temperature (°C)	Lognormal Distribution {GM, GSD, min, max}	Original AER Study Data Used
Fall River, MA	< 10	{0.711, 2.108, 0.1, 10}	New York, NY
	10 - 25	{1.139, 2.677, 0.1, 10}	
	> 25	{1.244, 2.177, 0.1, 10}	
Indianapolis, IN	< 10	{0.744, 1.982, 0.1, 10}	Detroit, MI and New York, NY
	10 - 20	{0.811, 2.653, 0.1, 10}	
	20 - 25	{0.785, 2.817, 0.1, 10}	
	> 25	{0.916, 2.671, 0.1, 10}	
Tulsa, OK	< 20	{0.407, 2.113, 0.1, 10}	Houston, TX
	20 - 25	{0.467, 1.938, 0.1, 10}	
	25 - 30	{0.422, 2.258, 0.1, 10}	
	> 30	{0.499, 1.717, 0.1, 10}	

Table 4-8. AERs for indoor residential microenvironments (ME-1) without A/C by study area and temperature.

Study Area	Daily Mean Temperature (°C)	Lognormal Distribution {GM, GSD, min, max}	Original AER Study Data Used
Fall River, MA	< 10	{1.016, 2.138, 0.1, 10}	New York, NY
	10 - 20	{0.791, 2.042, 0.1, 10}	
	> 20	{1.606, 2.119, 0.1, 10}	
Indianapolis, IN	< 0	{1.074, 1.772, 0.1, 10}	Detroit, MI and New York, NY
	0 - 10	{0.760, 1.747, 0.1, 10}	
	10 - 20	{1.447, 2.950, 0.1, 10}	
	20 - 25	{1.531, 2.472, 0.1, 10}	
	> 25	{1.901, 2.524, 0.1, 10}	
Tulsa, OK	< 10	{0.656, 1.679, 0.1, 10}	Houston, TX
	10 - 20	{0.625, 2.916, 0.1, 10}	
	> 20	{0.916, 2.451, 0.1, 10}	

4.4.2 Air Conditioning Prevalence for Indoor Residential Microenvironments

The selection of an AER distribution is dependent on the presence or absence of A/C. We assigned this housing attribute to indoor residential microenvironments using A/C prevalence data from the 2013 American Housing Survey (AHS).[39] A/C prevalence (specified in terms of does or does not have mechanical air conditioning) is distinct from usage (i.e., mechanical air conditioning is on or off), the latter ultimately represented by values selected from the AER distribution (relatively lower values would be associated with greater recirculation of indoor air) and dependent on the daily temperature. The A/C prevalence data were assigned to our study areas where the AHS data best matched our exposure simulation years (Table 4-9). In all three study areas, the sum of room unit and central A/C prevalence was used.

Table 4-9. American Housing Survey A/C prevalence from 2013 Current Housing Reports for selected urban areas.

Study Area [a]	Total Occupied Housing Units (x1000)	Number of Occupied Housing Units (x1000)					% of Occupied Housing Units		
		Central A/C	>1 Central A/C	1 Room Unit	2 Room Units	3+ Room Units	Central A/C	Window Units	Central & Window A/C
Fall River, MA	780.3	296.6	20.1	129.6	131.0	146.0	38	52	90
Indianapolis, IN	359.7	319.3	21.5	11.9	14.7	8.4	89	10	99
Tulsa, OK	262.0	233.3	7.1	12.1	6.9	61.2	89	10	99

[a] Data used were from the 2013 Metropolitan Area using a geography filter of 'not in central cities'. Because there were no data for the study areas data reported for nearby cites was used as follows: Fall River, MA - Boston, MA; Indianapolis - Louisville, KY; Tulsa, OK – Oklahoma City OK.

[39] Available at https://www.census.gov/programs-surveys/ahs/data/interactive/ahstablecreator.html.

4.4.3 AER Distributions for All Other Indoor Microenvironments

To estimate AER distributions for all non-residential, indoor environments (e.g., offices, libraries, schools, etc.), we relied on data generated as part of the U.S. EPA Building Assessment Survey and Evaluation (BASE) study (Persily and Gorfain, 2004; Persily et al., 2005), as was also done for the 2009 REA and REAs for other recent NAAQS reviews (e.g., U.S. EPA, 2014). In the BASE study, a total of 390 AER measurements were collected from 96 randomly selected office buildings throughout the U.S. using two methods, a volumetric and a carbon dioxide ratio method. In the vast majority of cases, the reported best estimate was generated using the volumetric method. The AER values for each office space were averaged, rather than using the individual measurements, because of the limited degree of variability in AER measurements for the same office space over a relatively short sampling period. We fitted exponential, lognormal, normal, and Weibull distributions to the 96 office space average AER values, and the best fitting of these was the lognormal. The fitted parameters for this distribution are a geometric mean of 1.109, geometric standard deviation of 3.015, and bounded by the lower and upper values of the sample data set {0.07, 13.8}.

4.4.4 Removal Rate for Indoor Microenvironments

To estimate pollutant removal rates from air within indoor microenvironments, we first evaluated the removal rates that had been estimated in the 2009 REA using data collected by Grontoft and Raychaudhuri (2004) on SO_2 deposition to a variety of building material surfaces under differing conditions of relative humidity. In the 2009 REA, this information was used to derive estimates for five indoor microenvironments: residences, office buildings, schools, restaurants, and other buildings (see 2009 REA Appendix B section 4.1). For the current REA, we simulated only two indoor microenvironments: residences and an aggregate ME representing all other indoor microenvironments. Therefore, we used the same removal rates that were derived for the 2009 REA for the residential ME and aggregated the estimated removal rates from the other four indoor MEs as follows. One thousand values were randomly sampled from the geometric means and standard deviations representing the removal rates for each of the four indoor MEs. Parameters describing a lognormal distribution for the new aggregate ME (for other indoor locations) were calculated using the 4,000 sampled values and are provided in Table 4-10.

Table 4-10. Parameter estimates of SO_2 removal rate distributions in two indoor microenvironments modeled by APEX.

Microenvironment	Removal (hr^{-1}) when Heating or Air Conditioning in Use				Removal (hr^{-1}) when Heating or Air Conditioning Not in Use			
	Geometric Mean	Standard Deviation	Lower Limit [a]	Upper Limit [a]	Geometric Mean	Standard Deviation	Lower Limit	Upper Limit
Indoor Residence	3.14	1.11	2.20	5.34	13.4	1.11	10.3	26.0
Indoor Other	3.32	1.37	1.53	5.07	N/A	N/A	N/A	N/A

[a] Lower and Upper Limits were approximated by the 10th and 90th percentile values.
[b] N/A not applicable, assumed to always have mechanical building ventilation in operation.
[c] From Table B.4-6 of 2009 REA.
[d] Derived from 4,000 values sampled from removal distributions representing four indoor microenvironments (Table B.4-6 of 2009 REA).

4.4.5 Factor for Estimating In-Vehicle/Near-Road Microenvironmental Concentrations

As was the case for the 2009 REA, there are no SO_2 measurement data available to develop a factor for estimating SO_2 concentrations inside vehicles resulting from the ambient air pollutant entering the microenvironment (and termed *PE factor*). The ratio of inside-vehicle ME concentrations to outdoor concentrations is commonly used to develop this PE factor. Thus, based on the outdoor concentration, one can estimate the inside-vehicle concentrations. Therefore, as was done for the 2009 REA, the PE factors used were developed from NO_2 data provided in Chan and Chung (2003) and used in the 2008 NO_2 REA (U.S. EPA, 2008a). As both SO_2 and NO_2 are gaseous, and data for PE factors are not broadly available for other gases, this was concluded to be a reasonable approach.

We note that pollutant removal rates inside vehicles might be different because SO_2 is more water soluble than NO_2, although we could not find removal rate data specific to motor vehicles. A comparison of indoor residential removal rates used for NO_2 (2008 NO_2 REA) to that of the 2009 SO_2 REA suggests that there might be greater removal of SO_2 within indoor microenvironments, indicating that use of the same PE factor for SO_2 as was used for NO_2 could lead to overestimation of inside-vehicle SO_2 concentrations.[40] Further, however, the in-vehicle NO_2 measurements on which the in-vehicle-to-outdoor-ratios were based might have included a small amount of in-vehicle emissions, potentially yielding a discrepancy between effective PE factors for NO_2 and SO_2. The additional uncertainty from this influential factor is expected to be

[40] NO_2 removal rates for the 2010 REA were assumed to range from 1.02 to 1.45 h^{-1}, based on six measurements obtained from a single house provided by Spicer et al. (1993). SO_2 removal rates for the 2009 REA were approximated by using SO_2 deposition collected data Grontoft and Raychaudhuri (2004) for a variety of building material surfaces under differing conditions of relative humidity and configured to five indoor microenvironments. The lower and upper limits of the removal rates ranged from 1.64 to 5.34 h^{-1}.

small compared to the overall uncertainty implied by using a uniform distribution that assumes all factors that influence variability and that are not directly accounted for have the same impact.

Chan and Chung (2003) measured inside-vehicle and outdoor NO_2 concentrations for three ventilation conditions: air-recirculation, fresh air intake, and with windows open. Mean in-vehicle-to-outdoor ratio values ranged from about 0.6 to just over 1.0, with higher values associated with increased ventilation (i.e., window open). A uniform distribution U{0.6, 1.0} was selected for the PE factor due to the simplified manner it is applied in this REA. For example, we could not consider influential characteristics such as use of vehicle ventilation systems due to the lack of data available to reasonably assign values for each study area.

4.5 ESTIMATING EXPOSURE

Based on the event-specific exposures estimated for each individual as described in the preceding sections, APEX identifies the occurrence of daily maximum 5-minute SO_2 exposures at or above specific levels, while at or above the target ventilation rate (i.e., an EVR ≥ 22 L/min-m^2). More specifically, this is the count of individuals (with asthma) experiencing a specific number of days per year (e.g., one or more, two or more, etc.) with exposures at or above specified 5-minute SO_2 concentrations (i.e., falling within bins representing different magnitudes of exposure) while at elevated ventilation.

The daily maximum 5-minute exposure concentrations (of people with asthma at elevated ventilation) are binned considering the overall features expected for the distribution of ambient SO_2 concentrations and population-based SO_2 exposures. Observed ambient concentrations are generally lognormally distributed – on average, 1-hour daily maximum concentrations are about 5 ppb, 90th percentile 1-hour daily maximum concentrations are typically below 20 ppb, while 99th and maximum 1-hour daily maximum concentrations can be a factor of 10 to 20 times higher than the mean (ISA, Table 2-13). It follows that because of this distribution of ambient air concentrations, it is likely that most simulated individuals will experience low daily maximum exposures (between 5 and 20 ppb), some will experience a daily maximum exposure between 20 and 100 ppb, while few will experience exposures above 100 ppb. Considering this and the relationship documented in the controlled human exposure studies between exposure concentration and percent of individuals estimated to experience a lung function decrement (section 4.6.2), exposure bins were as follows.

For exposure concentrations below 150 ppb, the exposure bins are set at 10 ppb increments (e.g., 10–20 ppb, 20–30 ppb, etc.); exposure concentrations at or above 150 and below 250 ppb are at 20 ppb increments though also including a bin for 200 ppb; and exposure concentrations at or above 250 are at 50 ppb increments, totaling 29 exposure bins. The smaller bin increments are used for lower exposure concentrations given the relatively greater number of exposure events expected to occur in that range and the desire to reduce potential for overestimation through use of larger size bins (see REA Planning Document, section 4.2.4.2). From this we summarize the number of days with maximum exposures within each exposure bin, such that the exposure model outputs are summarized as (1) counts of people exposed at least one day per year to a range of short-term peak SO_2 concentrations while at or above the target exertion level, and (2) counts of people experiencing multiple days per year with the maximum 5-minute exposure at or above a particular level while at or above the target exertion level.

4.6 RISK METRICS

Using the population exposure estimates, we derived two types of metrics to characterize potential population health risk: (1) comparison to benchmark concentrations; and, (2) lung function risk. As in the last review, these approaches are based on the body of evidence from the controlled human exposure studies reporting lung function decrements (as measured by changes in sRaw), as well as changes in other measures of lung function, respiratory symptoms, and various markers of inflammation, in adult study subjects having asthma. For both approaches, estimates are developed for two groups of individuals with asthma living in the study areas: adults with asthma (individuals older than 18 years), and children with asthma (individuals aged 5 to 18 years).

4.6.1 Comparison to Benchmark Concentrations

One of the two types of risk metrics in this assessment is based on comparison of estimated 5-minute exposures experienced while at an elevated ventilation rate to benchmark concentrations based on the controlled human exposure studies. In addition to its use in the 2009 SO_2 REA, the benchmark approach was used in past NO_2 and O_3 REAs (e.g., U.S. EPA, 2014), although ventilation rate does not play a role in the approach for the NO_2 REA. For this metric, the time-series of exposures for each APEX-simulated individual is used to identify the daily maximum 5-minute SO_2 concentrations that occur while at moderate or greater exertion. Based on all of the instances a daily maximum 5-minute exposure (while at or above the target EVR) is at or above a benchmark concentration, summaries of the individual-level exposures are produced and combined to generate a statistic for the simulated at-risk population in each study area. This statistic indicates the number (and percent) of simulated persons experiencing exposures at or above the benchmark concentrations, while at moderate or greater exertion.[41]

[41] A 'person-day' metric can be generated, indicating the total number of exceedances across the study area as a whole, but this metric is less informative for this review. The metric conflates variability in individual exposures, which can vary widely depending on the occurrence of peak concentrations and the distribution of time spent outdoors, and from a physiological perspective, creates an uninterpretable aggregate population exposure metric.

As in the 2009 REA, we have identified a set of benchmark concentrations to represent "exposures of potential concern" (75 FR 35527, June 22, 2010), 5-minute exposure concentrations for which there is potential for a respiratory response indicative of some level of bronchoconstriction to occur in an exposed individual, with the potential and the severity varying with the magnitude of the benchmark concentration. These levels are derived solely from the controlled human exposure studies, which can examine the health effects of SO_2 in the absence of copollutants that typically can confound results in epidemiologic analyses; thus, health effects observed in such controlled studies can confidently be attributed to a defined SO_2 exposure level.

Considering this information on variation in SO_2 exposures and severity of respiratory response, as described in the ISA and summarized in section 2.2.3 of the REA Planning Document, we concluded that it is appropriate, as in the last review, to use four benchmark concentrations: 100, 200, 300 and 400 ppb. As recognized in the last review, we consider exposures with respect to the 200 and 400 ppb 5-minute benchmark concentrations to be of particular interest because: (1) 400 ppb represents the lowest exposure concentration in controlled human exposure studies where moderate or greater lung function decrements occurred that were often statistically significant at the group mean level and frequently accompanied by respiratory symptoms; and (2) 200 ppb is the lowest exposure concentration in controlled human exposure studies at which moderate or greater lung function decrements were found in some individuals, although these lung function changes were not statistically significant when evaluated at the group mean level (75 FR 35527, June 22, 2010) (Table 4-11). Additionally, analyses of pooled datasets for study subjects with asthma that are responsive to SO_2 at concentrations below 1000 ppb found statistically significant increases in lung function decrements at 300 ppb (ISA, p. 5-19 to 5-20, 5-153; Johns et al., 2010). The lowest benchmark concentration (100 ppb) is one half the lowest exposure concentration tested by studies in which the exposure conditions allowed the study subjects to breathe freely.[42] We have included this benchmark concentration in consideration of the nonzero, albeit low (fewer than 10%), percentage of subjects with asthma experiencing moderate decrements in lung function at the 200 ppb exposure concentration and the lack of specific study data for some groups of individuals with asthma, such as primary-school-age children (ages 5 to 11) and those with severe asthma.[43]

[42] Studies of free-breathing subjects generally make use of small rooms in which the atmosphere is experimentally controlled such that study subjects are exposed by freely breathing the surrounding air (e.g., Linn et al., 1987).

[43] We have considered the evidence with regard to the response of individuals with severe asthma that are not generally represented in the full set of controlled human exposure studies. There is no evidence to indicate such individuals would experience moderate or greater lung function decrements at lower SO_2 exposure concentrations than individuals with moderate asthma. With regard to the severity of the response, the limited data that are

Table 4-11. Responses reported in controlled human exposure studies at a given benchmark concentration.

Benchmark Concentration (ppb)	Responses Reported in Controlled Human Exposure Studies [a]	
	Decrements in Lung Function	Respiratory Symptoms, Supporting Studies
400	Across studies of exposures at/above this concentration (400-500 ppb), 13-60% of exposed exercising study subjects with asthma experienced moderate decrements in lung function, and 4-40% experienced more severe responses [a b c]	"Stronger evidence, with some statistically significant increases in respiratory symptoms" (ISA, Table 5-2) [d]
300	Across studies of exposure at this concentration, 10-33% of exposed exercising study subjects with asthma experienced moderate decrements in lung function, and 0-40% experienced more severe responses [a e f]	"Limited evidence of SO_2-induced increases in respiratory symptoms in some people with asthma" (ISA, Table 5-2)
200	Across studies of exposures at this concentration, 7-9% of exposed exercising study subjects with asthma experienced moderate decrements in lung function, and up to 3% experienced more severe responses [a g]	
100	This is one half the lowest concentration tested in free-breathing exposure conditions [h]	

[a] Drawn from Table 5-2 of the ISA.

[b] Bronchoconstriction in individuals with asthma is the most sensitive indicator of SO_2-induced lung function effects and is characteristic of an asthma attack, and airway hyperresponsiveness (AHR) is a characteristic feature of individuals with asthma (ISA, section 5.2.1.2). As in the last review, the ISA describes as moderate decrements in lung function that involve at least a doubling in sRaw or at least a 15% reduction in FEV1; increases in sRaw of 200% or more and FEV1 reductions of 20% or more are indicated as more severe (ISA, section 1.6.1.1 and Table 5-2).

[c] Linn et al., 1983, 1987; Bethel et al.,1983; Roger et al., 1985; Magnussen et al., 1990; Horstman et al., 1986; ISA, Table 5-2.

[d] Lowest exposure finding both statistically significant lung decrements and respiratory symptoms (2008 ISA, section 3.1.3.1).

[e] Linn et al., 1988, 1990; ISA, Table 5-2.

[f] Statistically significant increases in lung function decrements in study subjects with asthma that are responsive to SO_2 at concentrations below 1000 ppb (ISA, pp. 5-19 to 5-20; Johns et al 2010).

[g] Linn et al., 1983, 1987; ISA, Table 5-2.

[h] Very limited data are available for this exposure concentration from five studies utilizing a mouthpiece to deliver pollutant concentrations (PA, section 3.2.1.3). In these studies, nasal absorption of SO_2 is bypassed during oral breathing, thus allowing a greater fraction of inhaled SO_2 to reach the tracheobronchial airways. As a result, individuals exposed to SO_2 through a mouthpiece are likely to experience greater respiratory effects from a comparable SO_2 exposure using a free breathing protocol (ISA, p. 5-23).Although few of these studies included an exposure to clean air while exercising that would have allowed for determining the effect of SO_2 *versus* that of exercise in causing bronchoconstriction, in those cases, the magnitudes of change in affected subjects appeared to be smaller than responses reported from studies at 200 ppb or more, with none indicating as much as a doubling in sRaw (PA, section 3.2.1.3).

available indicate a similar magnitude SO_2-specific response (in sRaw) as that for individuals with less severe asthma, although the individuals with more severe asthma are indicated to have a greater response to exercise prior to SO_2 exposure, indicating that those individuals "may have more limited reserve to deal with an insult compared with individuals with mild asthma" (ISA, p. 5-22).

4.6.2 Lung Function Risk

For lung function risk, we have focused on estimating the risk of experiencing SO_2-related increases in sRaw[44] that correspond to moderate decrements in lung function as described in the ISA.[45] The assessment estimates the number of people (and percent of the population) expected to experience such a decrement and the total number of occurrences of these effects per individual across the simulation period. Results include the number of people (and percent of population) estimated to experience at least one such decrement in a year and the number estimated to experience multiple decrements. Estimates are generated for each of two lung function response definitions: an increase in sRaw by at least 100% (Δ sRaw ≥ 100%), and an sRaw increase of at least 200% (Δ sRaw ≥ 200%). These measures of lung function risk are derived from the E-R function (discussed below) and the number of exposures (concomitant with moderate or greater exertion) among the population that are at or above each of a set of exposure concentrations estimated from the exposure modeling.

The E-R function is based on the controlled human exposure studies of decrements in lung function experienced by exercising individuals exposed to a range of 5-minute SO_2 concentrations. Table 4-12 presents all study summary data for changes in sRaw from all references from which individual study data are available (ISA, Table 5-2). Because the health response variable is binary, a generalized linear model (GLiM) was used to construct the E-R function (SAS, 2017), represented by the following

$$g(\mu) = \beta_0 + \beta_1 X \qquad \text{Equation 4-9}$$

Briefly explained, one important feature of GLiM is the function (g) used to link the structural component (i.e., the standard portion of a linear model, $\beta_0 + \beta_1 X$) to the mean of a conditional response distribution (μ). There are several types of link functions to use in fitting these regression models, the selection of which is generally guided by the empirical fit of the data, practical considerations, and knowledge of the form of the response distribution.

Two link functions (i.e., probit and logistic) were evaluated for developing the E-R function used in the 2009 REA (2009 REA, section 9.2 and Appendix C). Both functions are symmetrical, yielded similar model fits and had nearly similar functional shape, indicating either

---

[44] Although risk of lung function decrements in terms of both FEV_1 and sRaw were estimated in the REA for the last review, the risk related to increases in sRaw, a direct indicator of bronchoconstriction for which data are available across a more extensive set of exposure concentrations than FEV_1, was given greater emphasis and is the focus here.

[45] The ISA describes a doubling in sRaw (or a 15% reduction in FEV_1) to be a moderate lung function decrement (ISA, p. 1-17).

link function could be used to approximate risk. However, as is commonly observed for logistic functions, the lower and upper tails tend to be flatter when compared to a probit function, that is, the approach of the curve towards the horizontal axis is more gradual (see Figures 9-2 to 9-5 of 2009 REA). It followed that, when using the logistic function, the estimated health risk differed by a factor of 2 to 3 from that estimated using the probit function (see Tables 9-5 and 9-8 of 2009 REA). These differences in estimated risk were largely the result of combining the slightly higher probability of risk at low exposure concentrations when using the logistic function combined with the large number of simulated individuals having low 5-minute exposures; this is particularly the case for exposures less than 100 ppb (2009 REA, Figures 9-7 and 9-8). Thus, the probit model reduces the contribution of these exposures to risk estimates, for which E-R information is lacking, relative to that provide by the logistic function, thus better addressing the uncertainty in the E-R extrapolation to such low concentrations that appears magnified when using a logistic function. Further, as noted by the CASAC comments on a draft of the 2009 REA, assumptions regarding the distribution of individual thresholds for response support use of a probit function, which is based on the inverse of the cumulative standard normal distribution function, rather than a logistic function which assumes a logistic distribution, for estimating risk associated with population-based SO_2 exposures (Samet, 2009, pp. 14 and 60-63).

Based on the above factors, we used a probit model for this risk analysis as in the 2009 REA.[46] We used all of the data available[47] to fit the two separate E-R functions (for ΔsRaw $\geq$ 100% and ΔsRaw $\geq$ 200%), generating both the best fit regression as well as using variability associated with the predicted regression coefficients to provide lower and upper bounds of the risk estimation. To illustrate the E-R relationship indicated by these data, the percent of the study populations experiencing increases in sRaw is plotted in Figure 4-4. Further details regarding the E-R function, its application, and the interpretation of the estimated risk is provided below.

[46] The SAS procedure, PROC LOGISTIC, is used to fit the discrete response data by the method of maximum likelihood and using link=probit model option (SAS, 2017).

[47] As mentioned in the REA Planning Document, the concentration levels included in the regression can influence the model fit, in particular the area of particular interest in this REA (low concentration related predicted responses).

Table 4-12. Summary of controlled human exposure studies containing individual response data: number and percent of exercising individuals with asthma who experienced greater than or equal to a 100 or 200 percent increase in specific airway resistance (sRaw), adjusted for effects of exercise in clean air.

SO_2 (ppb)	Exposure duration (minutes)	N	Ventil-ation (l/min)	sRaw ≥100 (N)	sRaw ≥200 (N)	sRaw ≥100 (%)	sRaw ≥200 (%)	Reference
200	5	23	~48	2	0	8.7%	0.0%	Linn et al. (1983)[a]
200	10	40	~40	3	1	7.5%	2.5%	Linn et al. (1987)[b]
250	5	19	~50-60	6	3	31.6%	15.8%	Bethel et al. (1985)
250	5	9	~80-90	2	0	22.2%	0.0%	Bethel et al. (1985)
250	10	27	~42	0	0	0.0%	0.0%	Horstman et al. (1986)[a]
250	10	28	~40	1	0	3.6%	0.0%	Roger et al. (1985)
300	10	20	~50	2	1	10.0%	5.0%	Linn et al. (1988)
300	10	21	~50	7	2	33.3%	9.5%	Linn et al. (1990)
400	5	23	~48	3	1	13.0%	4.3%	Linn et al. (1983)[a]
400	10	40	~40	9.5	3.5	23.8%	8.8%	Linn et al. (1987)[b]
500	5	10	~50-60	6	4	60.0%	40.0%	Bethel et al. (1983)
500	10	27	~42	6	1	22.2%	3.7%	Horstman et al. (1986)[a]
500	10	28	~40	5	1	17.9%	3.6%	Roger et al. (1985)
600	5	23	~48	9	6	39.1%	26.1%	Linn et al. (1983)[a]
600	10	40	~40	13.5	9.5	33.8%	23.8%	Linn et al. (1987)[b]
600	10	20	~50	12	7	60.0%	35.0%	Linn et al. (1988)
600	10	21	~50	13	6	61.9%	28.6%	Linn et al. (1990)
1000	10	10	~40	6	2	60.0%	20.0%	Kehrl et al. (1987)
1000	10	28	~40	14	7	50.0%	25.0%	Roger et al. (1985)
1000	10	27	~42	15	7	55.6%	25.9%	Horstman et al. (1986)[a]

Data presented are from all studies from which individual data were available (ISA Table 5-2 and Figure 5-1) on percentage of individuals who experienced greater than or equal to a 100 or 200% increase in specific airway resistance (sRaw). Lung function decrements are adjusted for the effects of exercise in clean air (calculated as the difference between the percent change relative to baseline with exercise|SO_2 and the percent change relative to baseline with exercise|clean air).

[a] Data were not available for use in developing the E-R function for the 2009 SO_2 REA.

[b] Responses of mild and moderate asthmatics reported in Linn et al. (1987) are the average of the first and second round exposure responses following the first 10 min period of exercise.

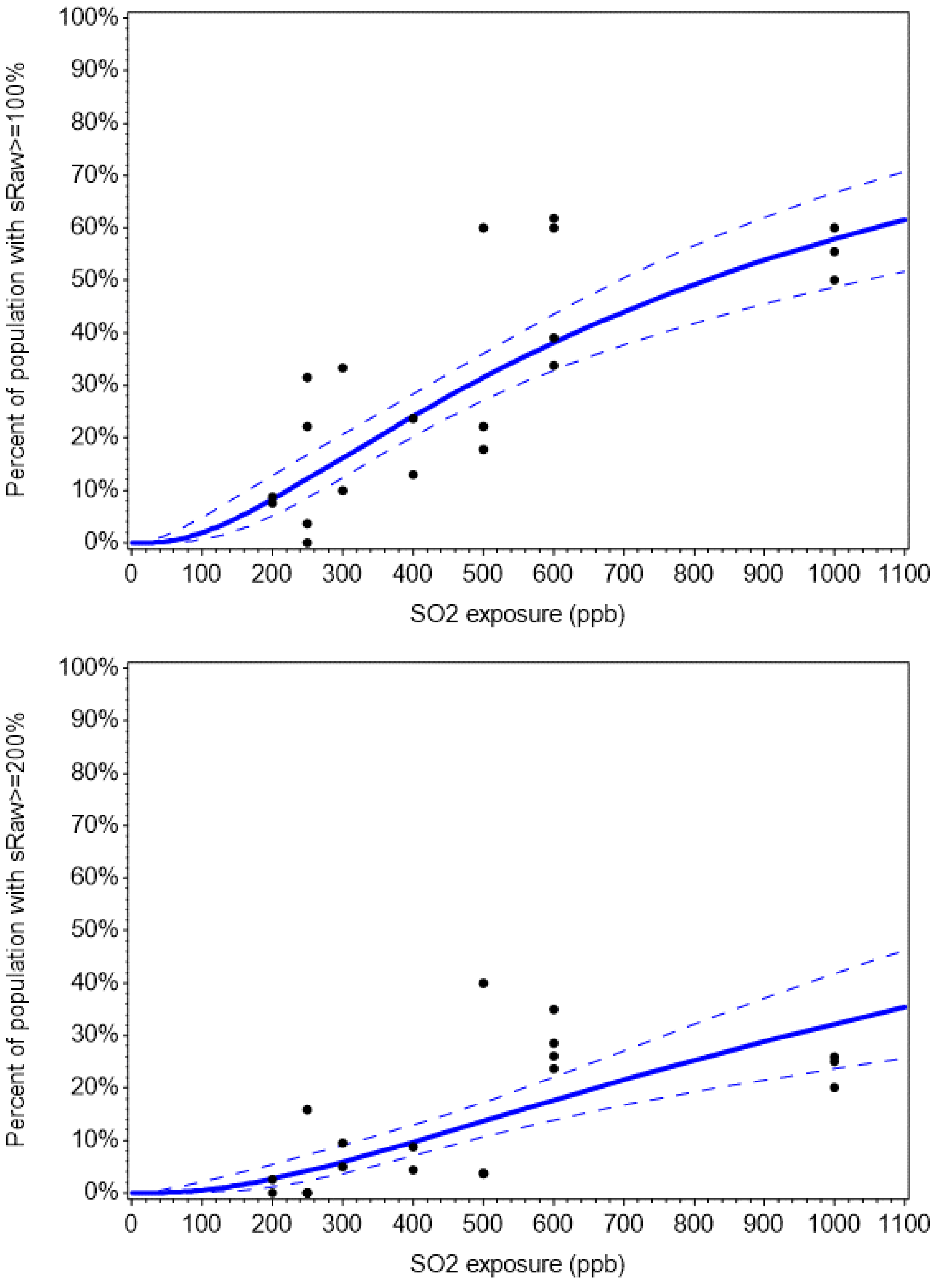


Figure 4-4. Percent of individuals experiencing changes in sRaw ≥ 100% (top panel) and sRaw ≥ 200% (bottom panel) using controlled human exposure study data (Table 4-12) fit using a probit regression (solid lines). Dashed lines indicate a 90 percent confidence interval for the mean response.

The intent of the REA approach described in this section is to calculate population risk, not individual risk. Thus, it is considered appropriate to focus on the mean response calculated from the limited number of subjects in the collection of independently performed controlled human exposure studies. Using the study subject data, we approximated a best fit function that represents a mean response for any daily maximum 5-minute exposure concentration and derived a confidence interval for it in order to present a range of estimated population response. We applied the best fit function to the exposures estimated for the entire simulated population, assuming the simulated at-risk population (people with asthma in the three study areas) is comprised of individuals that have a similar response frequency as the controlled human exposure study test subjects. The 90% confidence interval for the mean response was used to approximate lower and upper bounds of the E-R function and used estimate lower and upper bounds of the population risk as part of the uncertainty characterization (section 6.2.2.4). Given the objective of estimating risk associated with population exposures, this confidence interval is considered a reasonable approach for estimating the range of the estimated population risk.

An alternative approach for developing a range for the estimated risk might be a prediction interval that incorporates the spread of the individual study responses at each exposure concentration.[48] We concluded that such an approach – that can be biased by one particular study having responses outside of the curves representing the 90% confidence interval (e.g., see Figure 4-4) – would not provide an appropriate representation of population risk for sRaw responses. Given the wide-ranging responses in the individual studies, the very small number of subjects tested in each study leads us to conclude that such an approach that emphasized one or a few individual studies would be less likely to represent the response frequency for the entire population of people with asthma in each study area.

Using the exposure model counts of individuals with daily maximum 5-minute concentrations falling into the different bins (as described in section 4.5 above), the number of occurrences of lung function response is calculated by multiplying the number of exposures in an exposure bin by the response probability (given by the probit E-R function for the specified definition of lung function response) associated with the midpoint of that bin. Provided in Table 4-13 are single-year exposure estimates for children with asthma in the Fall River study area to demonstrate this calculation.

[48] In general, a prediction interval for a regression is useful in estimating a random future value of the dependent variable (y), while a confidence interval is useful in estimating the average (or expected) value of *y* variable given the same value of the independent variable (*x*).

Table 4-13. Example of risk calculation using estimated daily maximum 5-minute exposures of children with asthma in the Fall River study area.

Exposure Results		ER Function		Estimated Risk
5-minute SO_2 Exposure Bin [a] (ppb)	Number of Children [b] (n)	Bin Midpoint (ppb)	Calculated Response (fraction of population)	Number of Individuals Responding [c] (n)
0	44	5	2.49E-07	0
10	103	15	4.02E-05	0
20	149	25	2.92E-04	0
30	143	35	9.45E-04	0
40	190	45	2.12E-03	0
50	249	55	3.90E-03	0
60	345	65	6.28E-03	2
70	346	75	9.26E-03	3
80	477	85	1.28E-02	6
90	396	95	1.69E-02	6
100	379	105	2.15E-02	8
110	271	115	2.66E-02	7
120	196	125	3.21E-02	6
130	149	135	3.80E-02	5
140	70	145	4.41E-02	3
150	75	160	5.40E-02	4
170	36	180	6.80E-02	2
190	8	195	7.90E-02	0
200	5	205	8.65E-02	0
210	3	220	9.80E-02	0
230	0	240	1.14E-01	0
250	0	275	1.42E-01	0
300	0	325	1.82E-01	0
350	0	375	2.22E-01	0
400	0	425	2.60E-01	0
Total	3633			52

[a] The exposure bin includes daily maximum 5-minute exposures of at least that value, but less than that of the next exposure bin.
[b] This is the number of children with asthma experiencing the exposure while at moderate or greater exertion. In the Fall River study area, the total population of children with asthma is 3,641.
[c] Multiplying number of children by the calculated response, then rounded down to the nearest integer, gives the number of individuals responding.

For example, the midpoint of the 10-20 ppb bin is 15 ppb (Table 4-13). The 15 ppb exposure bin contains a total of 103 individuals who experienced a daily maximum 5-minute concentration in the simulated year of at least 10 ppb, but less than 20 ppb. The frequency/probability obtained from the probit function at 15 ppb (i.e., 4.02 E-05) is then used to estimate the number of the 103 persons that respond. To avoid accounting for and summing (numerically calculated) fractions of people, all risk estimates obtained by combining the number of individuals with the percent responding within each bin (i.e., the count of individuals responding) are truncated at the integer level. Therefore, in the Table 4-13 example for the 10-20 ppb bin, the number of individuals estimated to experience a response (i.e., 0.004 persons) is zero. After calculating the number of whole individuals estimated to respond in each bin, these are summed to generate the total estimated population risk (i.e., 52 children with asthma). Thus, 1.4% of children with asthma (52 divided by 3641) are estimated to experience at least one day in the simulated year with an sRaw increase of 100% or more) as a result of their daily maximum 5-minute SO_2 exposure.

Additionally, the contribution to risk estimates from each exposure bin is developed based on the apportionment of the risk estimates to the exposure bins. In this example, nearly 90% of the estimated risk is attributed to 5-minute concentrations at or above 50 ppb and less than 150 ppb. No children were estimated to experience a response at a 5-minute concentration below 50 ppb.[49]

4.7 APPROACH FOR CHARACTERIZING UNCERTAINTY AND VARIABILITY

An important issue associated with any population exposure and risk assessment is the assessment of variability and characterization of uncertainty. Variability refers to the inherent heterogeneity in a population or variable of interest (e.g., residential air exchange rates). The degree of variability cannot be reduced through further research, only better characterized with additional measurement. Uncertainty refers to the lack of knowledge regarding the values of model input variables (i.e., parameter uncertainty), the physical systems or relationships used (i.e., use of input variables to estimate exposure or risk or model uncertainty), and in specifying the scenario that is consistent with purpose of the assessment (i.e., scenario uncertainty). Uncertainty is, ideally, reduced to the maximum extent possible through improved measurement of key parameters and iterative model refinement. The following two sections describe the approaches we have used to assess variability (section 4.7.1) and to characterize uncertainty (section 4.7.2) in this REA. The primary outcome is a summary of variability and uncertainty evaluations conducted to date of our SO_2 exposure assessments and APEX exposure modeling, and the identification of the elements or areas of the assessment with which is associated the greatest uncertainty.

[49] Thus, it can be observed that in this example for this study area and air quality scenario, 50 ppb represents a limiting value for the response function among the lowest exposure level bins. Such values would be expected to differ with population and air quality scenario characteristics.

4.7.1 Assessment of Variability and Co-variability

The goal in addressing variability in the REA is to ensure that the estimates of exposure and risk reflect the variability of SO_2 concentrations in ambient air, population characteristics, associated SO_2 exposures, physiological characteristics of simulated individuals, and potential health risk across the study areas and for the simulated at-risk populations. In the REA, there are several algorithms that are used to account for variability when generating the two risk metrics. For example, variability may arise from differences in the population residing within census tracts (e.g., age distribution) and the activities that may affect population exposure to SO_2 (e.g., time spent outdoors, performing moderate or greater exertion level activities outdoors). The range of exposure and associated risk estimates are intended to reflect such sources of variability, although we note that the range of values obtained reflects the input parameters, algorithms, and modeling system used, and may not necessarily reflect the complete range of the true exposure or risk values.

We note also that correlations and non-linear relationships between variables input to the model can result in the model producing inaccurate results if the inherent relationships between these variables are not preserved. APEX is designed to account for co-variability, or linear and nonlinear correlation among the model inputs, provided that enough is known about these relationships to specify them. This is accomplished by providing inputs that enable the correlation to be modeled explicitly within APEX. For example, there is a non-linear relationship between the outdoor temperature and air exchange rate in homes. One factor that contributes to this non-linear relationship is that windows tend to be closed more often when temperatures are at either low or high extremes than when temperatures are moderate. This relationship is explicitly modeled in APEX by specifying different probability distributions of air exchange rates for different ambient air temperatures.

Important sources of the variability and co-variability accounted for by APEX and used for this SO_2 exposure analysis have been identified and summarized in section 6.1. Where possible, we identified and incorporated the observed variability in input data sets rather than employing standard default assumptions and/or using point estimates to describe model inputs.

4.7.2 Characterization of Uncertainty

While it may be possible to capture a range of exposure or risk values by accounting for variability inherent to influential factors, the true exposure or risk for any given individual within a study area may be unknown, although it can be estimated. To characterize health risks, exposure and risk assessors commonly use an iterative process of gathering data, developing models, and estimating exposures and risks, given the goals of the assessment, scale of the assessment performed, and limitations of the input data available. However, significant uncertainty often remains and emphasis is then placed on characterizing the nature of that uncertainty and its impact on exposure and risk estimates.

In section 6.2, we have summarized the most important uncertainties potentially affecting the exposure estimates derived for this assessment. In so doing, we recognize that uncertainties associated with APEX exposure modeling are also characterized in the REAs conducted for recent reviews of the primary NAAQS for NO_2, carbon monoxide, and O_3, along with other pollutant-specific issues (U.S. EPA, 2008a, 2010, 2014). Conclusions drawn from each of these characterizations are considered in light of new information and of the approaches used in this REA. Additionally, the new evaluations performed in the current REA have been synthesized following the approach outlined by WHO (2008) and used to identify, evaluate, and prioritize the most important uncertainties relevant to the estimated exposure and risk outcomes. The characterization presented in section 6.2 uses a predominantly qualitative approach supplemented by various model sensitivity analyses and input data evaluations, all complementary to quantitative uncertainty characterizations conducted for the 2007 O_3 REA by Langstaff (2007).

The approach used for this REA varies from that described by WHO (2008) in that a greater focus has been placed on evaluating the direction and the magnitude[50] of the uncertainty. This refers to qualitatively rating how the source of uncertainty, in the presence of alternative information, may affect the estimated exposures and health risk results. Following the identification of key uncertainties, we have subjectively scaled the overall impact of the uncertainty by considering the relationship between the source of uncertainty and the exposure concentrations (e.g., low, moderate, or high potential impact). Also to the extent possible, we have included an assessment of the direction of influence, indicating how the source of uncertainty may be affecting exposure or risk estimates (e.g., the uncertainty could lead to over- or under-estimates). Further, and consistent with the WHO (2008) guidance, section 6.2 discusses the uncertainty in the knowledge base (e.g., the accuracy of the data used, acknowledgement of data gaps) and, where possible, particular assessment design decisions (e.g., selection of particular model forms). The output of the uncertainty characterization is the summary in section 6.2 that describes, for each identified source of uncertainty, the magnitude of the impact and the direction of influence the uncertainty may have on the exposure and risk characterization results.

[50] This is synonymous with the “level of uncertainty” discussed in WHO (2008), section 5.1.2.2.

REFERENCES

Ainsworth BE, Haskell WL, Herrmann SD, Meckes N, Bassett DR Jr, Tudor-Locke C, Greer JL, Vezina J, Whitt-Glover MC, Leon AS. (2011). 2011 Compendium of Physical Activities: a second update of codes and MET values. Med Sci Sports Exerc. 43(8):1575-81. doi: 10.1249/MSS.0b013e31821ece12. See also: https://sites.google.com/site/compendiumofphysicalactivities/references

Burmaster DE. (1998). LogNormal distributions for skin area as a function of body weight. Risk Analysis. 18(1):27-32.

Chan AT and Chung MW. (2003). Indoor-outdoor air quality relationships in vehicle: effect of driving environment and ventilation modes. Atmos Environ. 37:3795-3808.

Che WW, Frey HC, Lau AK. (2014). Assessment of the effect of population and diary sampling methods on estimation of school-age children exposure to fine particles. Risk Analysis. 34(12):2066-2079.

Cohen J, Mallya H, Rosenbaum A. (2012). Updated Analysis of Air Exchange Rate, ICF International Memo to John Langstaff. Available in Appendix 5E of O_3 HREA.

Esmail S, Bhambhani Y, and Brintnell S. (1995). Gender differences in work performance on the Baltimore therapeutic equipment work simulator. Amer J Occup Therapy. 49:405-411.

Geyh AS, Xue J, Özkaynak H, Spengler JD. (2000). The Harvard Southern California chronic ozone exposure study: assessing ozone exposure of grade-school-age children in two Southern California communities. Environmental Health Perspectives. 108:265-270.

Glen G, Smith L, Isaacs K, McCurdy T, Langstaff J. (2008). A new method of longitudinal diary assembly for human exposure modeling. J Expo Sci Environ Epidemiol. 18:299-311.

Ford ES, Heath GW, Mannino DM, Redd SC. (2003). Leisure-time physical activity patterns among U.S. adults with asthma. Chest. 124:432-437.

Graham SE and McCurdy T. (2004). Developing meaningful cohorts for human exposure models. J Expos Anal Environ Epidemiol. 14(1):23-43.

Graham SE and McCurdy T. (2005). Revised Ventilation Rate (VE) Equations for Use in Inhalation-Oriented Exposure Models. Appendix A and D of EPA/600/R-06/129F. Available at: http://cfpub.epa.gov/ncea/cfm/recordisplay.cfm?deid=202543

Grontoft T and Raychaudhuri MR. (2004). Compilation of tables of surface deposition velocities for O_3, NO_2 and SO_2 to a range of indoor surfaces. Atmos Environ. 38:533-544.

Henry CJK. (2005). Basal metabolic rate studies in humans: measurement and development of new equations. Public Health Nutrition. 8(7A):1133-1152.

IOM. (2005). Dietary Reference Intakes for Energy, Carbohydrate, Fiber, Fat, Fatty Acids, Cholesterol, Protein and Amino Acids. Food and Nutrition Board. U.S. Institute of Medicine. National Academies Press. Available at: http://www.nationalacademies.org/hmd/Activities/Nutrition/SummaryDRIs/DRI-Tables.aspx. Table: Doubly Labeled Water Data Set: 10_DLW_Database.xls (downloaded 8/24/16)

Johnson T. (2002). A Guide to Selected Algorithms, Distributions, and Databases Used in Exposure Models Developed by the Office of Air Quality Planning and Standards. Revised Draft. Prepared for U.S. Environmental Protection Agency. Available at: https://www.epa.gov/sites/production/files/2013-08/documents/report052202.pdf

Isaacs K and Smith L. (2005). New Values for Physiological Parameters for the Exposure Model Input File Physiology.txt. Memorandum submitted to the U.S. Environmental Protection Agency under EPA Contract EP-D-05-065. NERL WA 10. Alion Science and Technology. Available in the 2009 SO_2 REA, Appendix B at: https://www3.epa.gov/ttn/naaqs/standards/so2/s_so2_cr_rea.html

Langstaff JE. (2007). Memorandum to Ozone NAAQS Review Docket (EPA-HQ-OAR-2005-0172). Analysis of Uncertainty in Ozone Population Exposure Modeling. Docket ID No. EPA-HQ-OAR-2005-0172-0174.

Linn W, Avol E, Peng R-C, Shamoo D, Hackney J. (1987). Replicated dose-response study of sulfur dioxide effects in normal, atopic, and asthmatic volunteers. American Review of Respiratory Disease. 136:1127-1134.

McCurdy T, Glen G, Smith L, Lakkadi Y. (2000). The National Exposure Research Laboratory's Consolidated Human Activity Database. J Expo Anal Environ Epidemiol. 10:566-578.

Persily A and Gorfain J. (2004). Analysis of Ventilation Data from the U.S. Environmental Protection Agency Building Assessment Survey and Evaluation (BASE) Study. National Institute of Standards and Technology, NISTIR 7145, December 2004.

Persily A, Gorfain J, Brunner G. (2005). Ventilation Design and Performance in U.S. Office Buildings. ASHRAE Journal. April 2005, 30-35.

Samet J. (2009). Letter from Johnathan M. Samet, Chair, Clean Air Scientific Advisory Committee, to Administrator Lisa P. Jackson. Re: Clean Air Scientific Advisory Committee’s (CASAC) Review of EPA's Risk and Exposure Assessment (REA) to Support the Review of the SO_2 Primary National Ambient Air Quality Standards: Second Draft. May 18, 2009.

Santuz P, Baraldi E, Filippone M, Zacchello F. (1997). Exercise performance in children with asthma: is it different from that of healthy controls? Eur Respir J. 10:1254-1260.

SAS. (2017). Base SAS® 9.4 Procedures Guide, Seventh Edition. Available at: http://documentation.sas.com/api/collections/pgmmvacdc/9.4/docsets/proc/content/proc.pdf?locale=en#nameddest=bookinfo. Also used were SAS/ETS 14.2 User's Guide, edited by A. Baxter, E. Huddleston and SAS/QC 14.2 User's Guide, edited by A. Baxter, V. Clark, E. Huddleston, S. Prabhu, R. Rodriguez, D. Sawyer, J. Simmons.

Schofield WN. (1985). Predicting basal metabolic rate, new standards, and review of previous work. Hum Nutr Clin Nutr. 39C(S1):5-41.

Spicer CW, Kenny DV, Ward GF, Billick IH. (1993). Transformations, lifetimes, and sources of NO_2, HONO, and HNO3 in indoor environments. JAWMA. 43(11):1479-1485.

U.S. DHHS. (1999). Promoting physical activity: a guide for community action. Department of Health and Human Services, Centers for Disease Control and Prevention (see Table 2-1). Available at: http://www.cdc.gov/nccdphp/dnpa/physical/pdf/PA_Intensity_table_2_1.pdf

U.S. EPA. (1989). Exposure Factors handbook. Office of Health and Environmental Assessment, Washington, DC, EPA/600/8-89/043.

U.S. EPA. (1997). Exposure Factors Handbook (Final Report, 1997). Office of Research and Development, National Center for Environmental Assessment, Washington, DC, EPA/600/P-95/002F a-c. Available at: https://cfpub.epa.gov/ncea/risk/recordisplay.cfm?deid=12464

U.S. EPA. (2007). Ozone Population Exposure Analysis for Selected Urban Areas. Office of Air Quality Planning and Standards, Research Triangle Park, NC, EPA-452-R-07-010, July 2007. Available at: http://www.epa.gov/ttn/naaqs/standards/ozone/s_o3_cr_td.html

U.S. EPA. (2008a). Risk and Exposure Assessment to Support the Review of the NO_2 Primary National Ambient Air Quality Standard. Office of Air Quality Planning and Standards, Research Triangle Park, NC, EPA-452/R-08-008a, November 2008. Available at: https://www3.epa.gov/ttn/naaqs/standards/nox/s_nox_cr_rea.html

U.S. EPA. (2008b). Integrated Science Assessment (ISA) for Sulfur Oxides – Health Criteria (Final Report). National Center for Environmental Assessment-RTP Division, Office of Research and Development, Research Triangle Park, NC, EPA-600/R-08/047F, September 2008. Available at: http://cfpub.epa.gov/ncea/cfm/recordisplay.cfm?deid=198843

U.S. EPA. (2009). Risk and Exposure Assessment to Support the Review of the SO_2 Primary National Ambient Air Quality Standard. Office of Air Quality Planning and Standards, Research Triangle Park, NC, EPA-452/R-09-007, July 2009. Available at: https://www3.epa.gov/ttn/naaqs/standards/so2/data/200908SO2REAFinalReport.pdf

U.S. EPA. (2010). Quantitative Risk and Exposure Assessment for Carbon Monoxide - Amended. Office of Air Quality Planning and Standards, Research Triangle Park, NC, EPA-452/R-10-006, July 2010. Available at: https://www.epa.gov/naaqs/carbon-monoxide-co-standards-risk-and-exposure-assessments-current-review

U.S. EPA. (2014). Health Risk and Exposure Assessment for Ozone. Office of Air Quality Planning and Standards, Research Triangle Park, NC, EPA-452/R-14-004a, August 2014. Available at: https://www.epa.gov/naaqs/ozone-o3-standards-risk-and-exposure-assessments-current-review

U.S. EPA. (2017a). Air Pollutants Exposure Model Documentation (APEX, Version 5) Volume I: User's Guide. Office of Air Quality Planning and Standards, Research Triangle Park, NC, January 2017. EPA-452/R-17-001a. Available at: https://www.epa.gov/fera/apex-user-guides

U.S. EPA. (2017b). Air Pollutants Exposure Model Documentation (APEX, Version 5) Volume II: Technical Support Document. Office of Air Quality Planning and Standards, Research Triangle Park, NC, January 2017. EPA-452/R-17-001b. Available at: https://www.epa.gov/fera/apex-user-guides

U.S. EPA. (2017c). The Consolidated Human Activity Database – Master Version (CHAD-Master). Technical Memorandum. Office of Research and Development, National Exposure Research Laboratory, Research Triangle Park, NC, 27711. In preparation. Previous version (09/15/2014). Available at: https://www.epa.gov/healthresearch/consolidated-human-activity-database-chad-use-human-exposure-and-health-studies-and

U.S. EPA. (2017d). Integrated Science Assessment (ISA) for Sulfur Oxides – Health Criteria (Final). National Center for Environmental Assessment-RTP Division, Office of Research and Development, Research Triangle Park, NC, EPA/600/R-17/451, December 2017. Available at: https://cfpub.epa.gov/ncea/isa/recordisplay.cfm?deid=338596

van Gent R, van der Ent K, van Essen-Zandvliet LEM, Rovers MM, Kimpen JLL; de Meer G, Klijn PHC. (2007). No difference in physical activity in (Un)diagnosed asthma and healthy controls. Pediatric Pulmonology. 42:1018-1023.

Whitfield R, Biller W, Jusko M, Keisler J. (1996). A Probabilistic Assessment of Health Risks Associated with Short- and Long-Term Exposure to Tropospheric Ozone. Argonne, IL: Argonne National Laboratory.

WHO. (2008). WHO/IPCS Harmonization Project Document No. 6. Part 1: Guidance Document on Characterizing and Communicating Uncertainty in Exposure Assessment. International Programme on Chemical Safety, World Health Organization, Geneva, Switzerland. Available at: http://www.who.int/ipcs/methods/harmonization/areas/exposure/en/

Xue J, McCurdy T, Spengler J, and Özkaynak H. (2004). Understanding variability in time spent in selected locations for 7-12 year old children. J Expo Anal Environ Epidemiol. 14:222-233.

Zahran HS and Bailey C. (2013). Factors associated with asthma prevalence among racial and ethnic groups - United States, 2009-2010 Behavioral Risk Factor Surveillance System. Journal of Asthma. 50(6):583-589.

5 POPULATION EXPOSURE AND RISK RESULTS

Exposure and risk results are presented here for simulated populations residing in the three study areas – Fall River, MA, Indianapolis, IN, and Tulsa OK – for a three-year air quality scenario in which air quality conditions just meet the current primary SO_2 standard. The approaches used to link air quality modeling, ambient concentration measurements, exposure modeling, and controlled human exposure study data in this assessment are summarized in Figure 2-2. Briefly, and as described in more detail in chapter 3, first AERMOD predicts hourly SO_2 concentrations at air quality receptors within a spatial grid for each study area. Then, the complete annual temporal pattern of 5-minute continuous ambient monitor concentrations local to each study area was combined with the AERMOD-predicted 1-hour concentrations to generate 5-minute concentrations at every air quality receptor. As described in Chapter 4, APEX used the 5-minute air quality surface in each study area along with U.S. census block population demographics to estimate the number of days per year each simulated individual in a particular study area experiences a daily maximum 5-minute SO_2 exposure at or above 5-minute benchmark levels of 100, 200, 300, and 400 ppb. These short-term exposures were evaluated for children (5-18 years old) and adults (>18 years old) with asthma when the exposure corresponded with moderate or greater exertion (i.e., the individual's EVR was ≥22 L/minute-m^2). And finally, simulated individuals expected to experience a lung function decrement (i.e., doubling or larger increase in sRaw) were estimated by linking the population-based daily maximum 5-minute exposures with an exposure-response function derived from controlled human exposure study data (section 4.6.2)

Study area characteristics and the composition of the simulated population are provided in section 5.1. Exposure results are presented in a series of tables that allow for simultaneous comparison of the exposure and risk metrics across the three study areas and three simulation years. Two types of results are provided for each modeling domain: (1) the percent of the simulated subpopulation exposed at or above selected benchmarks, stratified by the number of occurrences (i.e., days) in a year (section 5.2) and (2) the percent of the simulated subpopulation experiencing a doubling or larger increase in sRaw, also stratified by the number of days in a year (section 5.3). Tables summarizing all of the exposure and risk results for each study area, exposure and response level, [1] and simulated at-risk population are provided in Appendix J.

[1] As described in section 4.5, exposure model output includes the number and percent of individuals at or above the benchmark levels and several other exposure levels used for estimating lung function risk.

5.1 CHARACTERISTICS OF THE SIMULATED POPULATION AND STUDY AREAS

The three study areas differ in population, geographic size, and demographic features (as summarized in Table 5-1 and Figures 5-1 through 5-3).[2] In each study area, APEX simulated SO_2 exposures for thousands of individuals,[3] the demographic features of which were based on the information associated with the thousands of census blocks within each area (as described in section 4.1 above).

Asthma prevalence in each modeling domain was estimated based on the NHIS asthma prevalence data and the demographic characteristics for each study area (e.g., age, sex and family income) using the methodology summarized in section 4.1.2. Accordingly, the percent of the simulated populations with asthma within the exposure modeling domain varied by study area (Table 5-1). The exposure modeling domain for Tulsa had the lowest percent of adults with asthma (7.2%), while Indianapolis had the lowest percent of children with asthma (9.7%). Fall River had the highest percent of children with asthma (11.2%), while Indianapolis had the highest percent of adults with asthma (8.3%). The statistics presented here are the aggregate of the study area as a whole, within which asthma prevalence varied widely as the modeling approach fully accounted for the variation in asthma prevalence across census blocks with demographic factors such as poverty, age, and sex (described in section 4.1.2).[4] Nationally, asthma prevalence is 7.8%; for children it is 8.4% and for adults it is 7.6% (PA, Table 3-2). The asthma prevalence for children, adults, and the total population estimated for each of the three study areas are all greater than that of the National asthma prevalence, except for adults in Tulsa which has a slightly lower asthma prevalence. This suggest that overall, the at-risk population simulated in the three study areas could represent at-risk populations in other U.S. area that have a similarly above average asthma prevalence.

---

[2] Specific census block (or tract) identifiers used for the simulations are documented in the APEX 'sites' files for these simulations.

[3] While precisely 30,000 children and 70,000 adults were simulated as part of each APEX model run, the number of individuals estimated to be exposed are appropriately weighted to reflect the actual population residing within the census blocks that comprise each respective study area.

[4] Representing the variation in asthma prevalence that occurs at the census block level provides a level of resolution for identification of at-risk individuals that is generally comparable with the resolution of the spatially variable ambient air concentrations at air quality receptors. In this way, the population in census blocks with higher-concentration air quality receptors is represented appropriately with regard to asthma prevalence and exposures of the at-risk individuals with asthma are not under-represented.

Table 5-1. Summary of study area features and the simulated population.

Study Area (# census tracts | # census blocks)	Population Group (age range)	Total Population	Population with Asthma	% of Population with Asthma
Fall River (56 | 4,364)	Children (5-18)	32,424	3,641	11.2 %
	Adults (19-95)	151,450	12,304	8.1 %
	All (5-95)	183,874	15,945	8.7 %
Indianapolis (172 | 12,310)	Children (5-18)	112,366	10,851	9.7 %
	Adults (19-95)	435,602	36,217	8.3 %
	All (5-95)	547,968	47,068	8.6 %
Tulsa (114 | 7,694)	Children (5-18)	49,482	5,484	11.1 %
	Adults (19-95)	207,941	15,049	7.2 %
	All (5-95)	257,423	20,533	8.0 %

There are also differences among the study areas with regard to the spatial distribution of the population (Figures 5-1 to 5-3).[5] In the Fall River study area (Figure 5-1), the most highly populated census tracts (6,000 to 9,000 people per tract) were generally toward the outer edges of the study area, with the exception of one highly populated tract encompassing the primary source. Most census tracts in the Fall River study area (86%) had a population of fewer than 6,000 people per tract, with a few tracts (23%) having fewer than 3,000 people per tract. In the Indianapolis study area (Figure 5-2), most tracts also had fewer than 6,000 people per tract (84%), though several tracts had greater than 9,000 people, one of which is located just south of the collection of modeled emission sources. The census tracts in the Tulsa study area were the least populated when compared to tract populations in the other two study areas, with all but one tract having fewer than 6,000 people and nearly 60% of tracts having fewer than 3,000 people per tract (Figure 5-3).

[5] Data used for these figures were obtained from https://www.census.gov/geo/maps-data/data/gazetteer2010.html. An identical scale was used for the three figures, progressing by increments of 3,000 people to allow for appropriate comparisons. For illustrative purposes, census tract population data were used for these maps to better view the overall population distribution across the study area rather than using census block level data because many of the smallest sized blocks were not viewable.

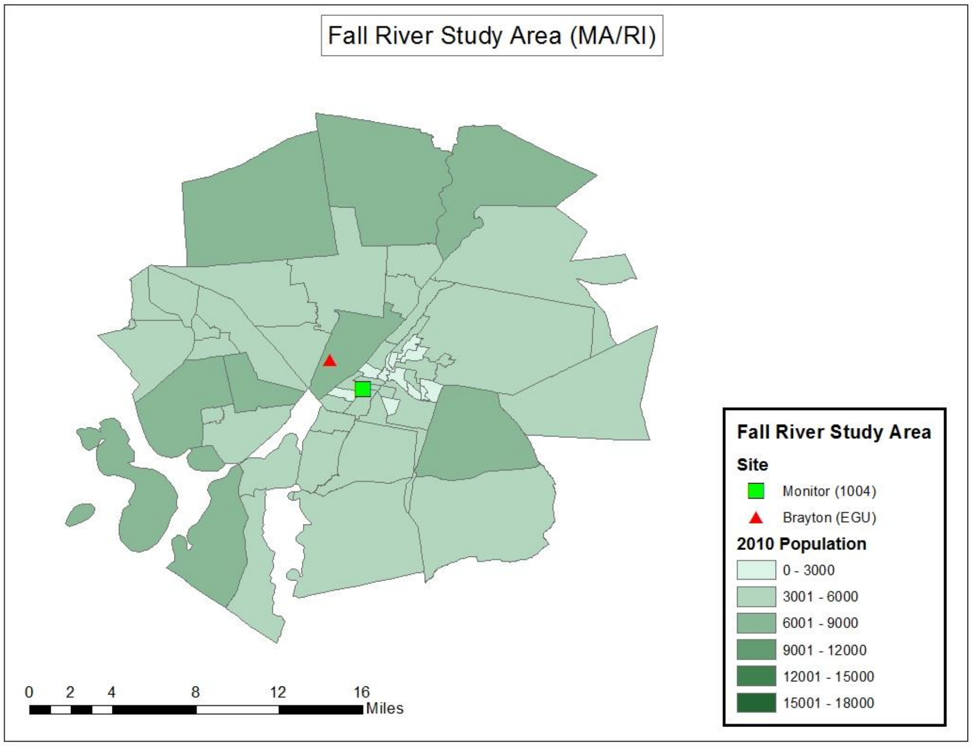


Figure 5-1. Population in the Fall River study area considering 2010 U.S. Census tracts.

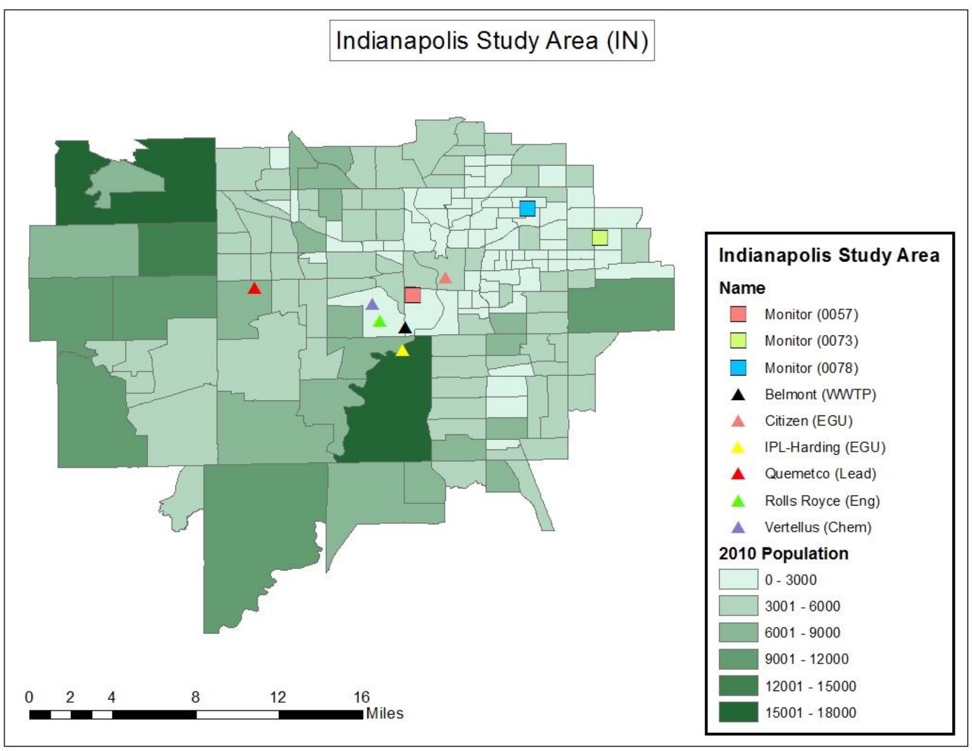


Figure 5-2. Population in the Indianapolis study area considering 2010 U.S. Census tracts.

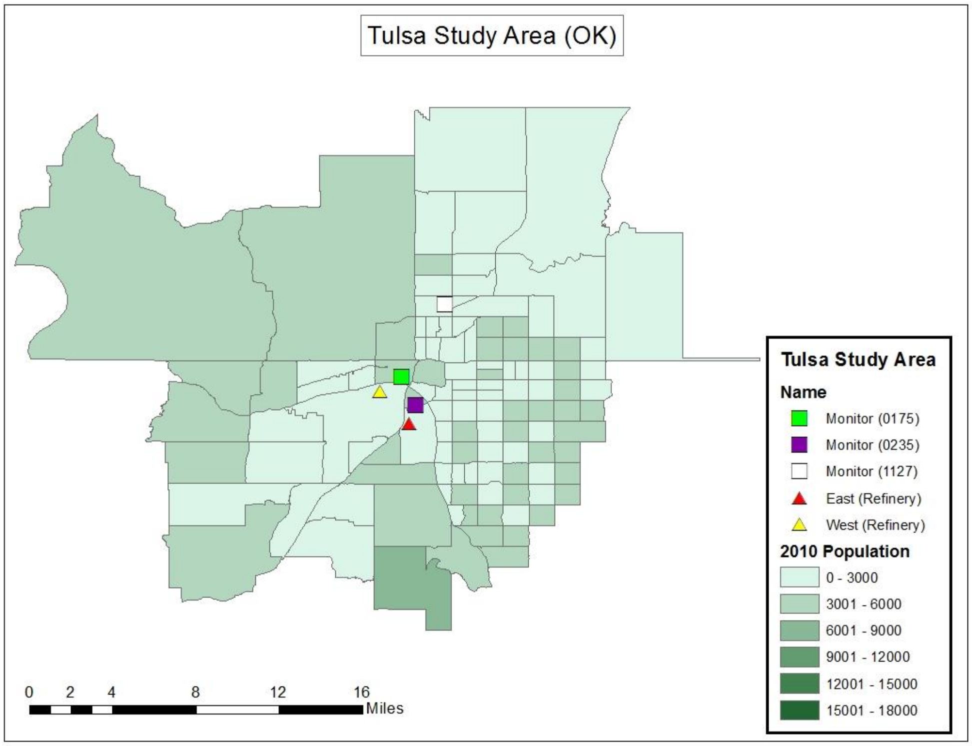


Figure 5-3. Population in the Tulsa study area considering 2010 U.S. Census tracts.

5.2 EXPOSURES AT OR ABOVE BENCHMARK CONCENTRATIONS

There were few simulated individuals estimated to experience 5-minute exposures at or above the three highest benchmark levels (200, 300, and 400 ppb), in any of the study areas (Tables 5-2 and 5-3). Regarding the two highest benchmarks of 300 ppb and 400 ppb, neither children nor adults with asthma had any 5-minute exposures at or above these levels in the Fall River and Tulsa study areas. In the Indianapolis study area, a small fraction (<1%) of the simulated population of children with asthma was estimated to experience exposures at or above 300 ppb and 400 ppb in the first year of the 3-year simulation. The relatively few exposures at or above 300 ppb is consistent with the limited number of occurrences of these high 5-minute concentrations in the air quality data set (Tables 3-14 to 3-16) for the air quality scenario modeled.[6] The next highest benchmark, 200 ppb, was also rarely exceeded, and when it was, a daily maximum 5-minute exposure only occurred at or above this level on no more than one day in the year and for a small fraction of the simulated at-risk population (<0.1% to 1.0%).

[6] Air quality was adjusted to just meet the existing standard of 75 ppb, as a 3-year average of 99th percentile annual daily maximum 1-hour concentrations.

Sensitivity analyses described in section 6.2.2 illustrate some variation from the estimates presented in this section. The use of alternative exposure model inputs, such as an alternative approach to adjust ambient concentrations to meet the existing standard and an alternative method to combine patterns of monitored 5-minute concentrations with modeled receptors can, in some instances, contribute to somewhat higher estimates. Overall, such differences based on the use of alternative approaches evaluated are not large and mainly affected estimated exposures at or above the 100 ppb benchmark (i.e., ranging from no difference to a few percentage points).

Given the findings noted above for the higher benchmark levels, discussion here of differences across air quality years and simulated populations focuses on the lowest benchmark level. Regarding this benchmark level (100 ppb), the Tulsa study area did not have more than 0.2% of either simulated at-risk population estimated to experience one or more days with a 5-minute exposure at or above 100 ppb (Tables 5-2 and Table 5-3). Thus, the discussion here focuses primarily on the Fall River and Indianapolis study area results.

Across the three years modeled, the highest population exposures were estimated in the first year. This is seen with the yearly estimates of the percent of the simulated populations expected to experience one, two or more days with exposures above benchmark levels (Tables 5-2 and Table 5-3). For example, in considering exposure results for children with asthma having at least one daily maximum 5-minute exposure at or above 100 ppb in Fall River during the first year, the percent was 32.7%, while air quality for the subsequent years yielded a lower percent (13.2% and 12.3%, respectively). Such year-to-year variability in the estimated exposures can be expected given variability in ambient concentrations across sequential years (e.g., Table 3-11 to Table 3-13), largely resulting from actual variability in emissions and meteorology in the air quality modeling.[7] Year-to-year variability was also observed for the Indianapolis results for 100 ppb, although the range (nine percentage points) was smaller.

Across the three areas, a greater proportion of simulated children with asthma were estimated to experience exposures at or above benchmark levels compared to adults with asthma. For example, for the three years in the Fall River study area, as many as 12.3 to 32.7% of children with asthma were estimated to experience at least one daily maximum 5-minute exposure at or above 100 ppb, while the range in the percent of adults with asthma exposed was from 1.3 to 5.1% (Table 5-2). The number of days per year with exposure above benchmarks was also greater for children with asthma compared to adults with asthma. For example, no

[7] The emissions for the main source in Fall River declined appreciably over the 3-year simulation period, also likely contributing to the variation observed in the annual exposure estimates. Note, the air quality adjustment used to create the hypothetical air quality scenario of conditions just meeting the existing standard (with its three-year form) maintains the year-to-year variability in emissions and meteorology, yielding high and low ambient concentration years within the 3-year period.

simulated adults with asthma in the Fall River study area were estimated to have more than three days in a year with a daily maximum 5-minute exposure at or above 100 ppb, while on average across the 3-year period, 0.9% of children with asthma were estimated to have four or more days at or above that benchmark (Table 5-3). Such differences between these two populations are expected given that higher exposures are more frequent outdoors (see sections 2.1.2 and below) and that children spend more time outdoors and at a greater frequency compared to adults.

Table 5-2. Percent and number of children and adults with asthma estimated to experience at least one day per year with a SO_2 exposure at or above 5-minute benchmark concentrations while breathing at elevated rate, air quality adjusted to just meet the existing standard.

Study area	Population group	Benchmark concentration (ppb)	Percent (and number) of population with asthma having at least one day per year when 5-minute SO_2 exposure $\geq$ benchmark			
			Year 1	Year 2	Year 3	Average
Fall River	children	100	32.7 (1,192)	13.2 (480)	12.3 (447)	19.4 (706)
		200	0.2 (8)	0 [a]	0	<0.1 [a] (3)
		300	No exposures at or above this benchmark			
	adults	100	5.1 (625)	1.9 (229)	1.3 (162)	2.8 (339)
		200	<0.1 (2)	0	0	<0.1 (1)
		300	No exposures at or above this benchmark			
Indianapolis	children	100	27.0 (2,932)	22.3 (2,419)	18.0 (1,947)	22.4 (2,433)
		200	1.0 (112)	0	0.9 (101)	0.7 (71)
		300	0.8 (89)	0	0	0.3 (30)
		400	0.3 (33)	0	0	0.1 (11)
	adults	100	4.3 (1,549)	3.8 1,369)	2.9 (1,051)	3.7 (1,323)
		200	0.1 (43)	0	0.2 (62)	0.1 (35)
		300	<0.1 (31)	0	0	<0.1 (10)
		400	<0.1 (24)	0	0	<0.1 (8)
Tulsa	children	100	0.2 (13)	0.2 (8)	<0.1 (1)	0.1 (7)
		200	No exposures at or above this benchmark			
	adults	100	0.1 (14)	<0.1 (8)	0	<0.1 (7)
		200	No exposures at or above this benchmark			

[a] < 0.1 represents nonzero estimates below 0.1%. A zero (0) indicates there were no individuals having the specified exposure.

Table 5-3. Percent of children and adults with asthma estimated to experience multiple days per year with a SO_2 exposure at or above 5-minute benchmark concentrations while breathing at elevated rate, air quality adjusted to just meet the existing standard.

Benchmark concentration (ppb)	Percent of population with asthma having multiple days per year when 5-minute SO_2 exposure ≥ benchmark [a]								
	Fall River			Indianapolis			Tulsa		
	≥2 days	≥4 days	≥6 days	≥2 days	≥4 days	≥6 days	≥2 days	≥4 days	≥6 days
	Children, aged 5 to 18 years								
100	5.5 (1.6 - 12.2)	0.9 (<0.1[b] - 2.6)	0.2 (0 - 0.6)	6.8 (4.7 - 8.0)	0.8 (0.3 - 1.0)	0.1 (<0.1 - 0.2)	0 [b]	0	0
200	no study area results included multiple days per year at or above this benchmark level								
	Adults, aged 19 to 95 years								
100	0.2 (<0.1 – 0.4)	0	0	0.5 (0.4 – 0.6)	<0.1 (0 - <0.1)	<0.1 (0 - <0.1)	0	0	0
200	no study area results included multiple days per year at or above this benchmark level								

[a] These estimates are summarized from the single year data provided in Appendix J. The first value in each cell is the average across the three years; the range is provided in parentheses.
[b] < 0.1 represents nonzero estimates below 0.1%. Zero (0) indicates there were no individuals having the specified exposure.

We also evaluated the microenvironments where the highest exposures occurred in the three study areas, as was done for the 2009 REA. With the summary information APEX provides for the simulated population is the total time spent in each microenvironment and for every exposure level across the entire simulation period.[8] [9] We summed the total time that simulated individuals in the population spent at or above each of the exposure levels and calculated the percent of this time occurring in each of the microenvironments (Figure 5-4). Consistent with findings from the 2009 REA, the majority (about 90% or more) of the time that the population is exposed at or above 100 ppb occurs in outdoor MEs for all three study areas.

[8] For all of these APEX simulations, children are the simulated population group, of which a subset are children with asthma. The APEX ME summary output is directly for that base population (i.e., all children) and cannot be edited to reflect a subset of that population (e.g., specific age groups of children). Because there are no modifications made to simulate children with asthma (i.e., all children use the same physiological and activity pattern data), inferences made regarding exposures for the total population of children in this analysis are applicable to the subset of simulated children with asthma.

[9] This default ME summary output summarizes all exposure time for the entire simulation, thus it reflects instances where individuals are at any exertion level (e.g., resting, vigorous, etc.). Nevertheless, the presentation here remains informative in this assessment, particularly considering that it is likely the vigorous exertion level activities are also linked to particular MEs such as those outdoors.

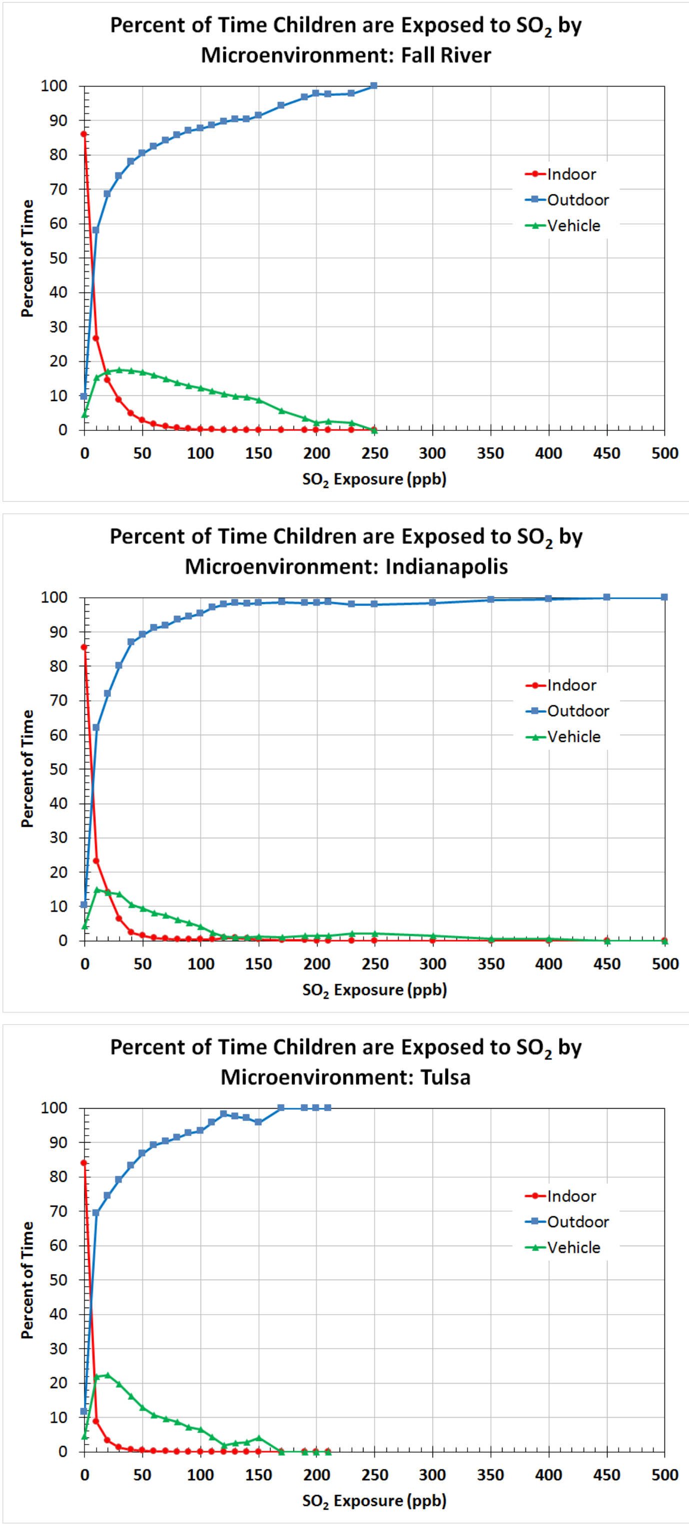


Figure 5-4. Percent of children's time in indoor, outdoor, and vehicle MEs while exposed to SO_2 in Fall River (top), Indianapolis (middle), and Tulsa study areas.

5.3 LUNG FUNCTION DECREMENTS ASSOCIATED WITH 5-MINUTE SO_2 EXPOSURES

There were few simulated individuals estimated to experience SO_2-related increases in sRaw of at least 100% in any of the three study areas under air quality conditions just meeting the existing standard (Tables 5-4 and 5-5). We additionally note that as mentioned above for the benchmark comparisons, sensitivity analyses using alternative exposure model inputs described in section 6.2.2 indicate that the percent of individuals estimated to experience lung function decrements of interest can vary from these estimates, although such differences are not large. Additionally, as discussed in section 5.4 below, differences among the three study areas with regard to the extent of areas within each study area with higher DVs and greater population appears to contribute to the relatively higher estimates for the Fall River and Indianapolis study areas. As recognized in the PA, such exposure circumstances are particularly informative to consideration of public health protection provided by the current SO_2 standard.

In the Fall River and Indianapolis study areas, on average across the three-year period, as many as 1.3% of children with asthma were estimated to experience at least one day per year with an SO_2-related increase in sRaw of 100% or more; in a single year, the percent is as high as 1.5% (Table 5-4). The percent of children with asthma estimated to experience two or more such days with an SO_2-related increase in sRaw of 100% or more ranged as high as 0.8% in a single year, while on average across the three years it was as high as 0.7% of children with asthma (Table 5-5). When considering SO_2-related increases in sRaw of 200% or more, on average as many as 0.3% of children with asthma were estimated to experience this lung function decrement. The percent of adults estimated to experience lung function decrements was lower than that of children, due to adults having a lesser amount of time spent outdoors and lower frequency of outdoor events, leading to lower exposures relative to those estimated for children.

Based on the design of the exposure assessment and how estimated exposures are summarized for the risk calculation (i.e., use of exposure concentration bins), the number of individuals falling within each exposure concentration bin is used to derive the number of individuals estimated to experience the lung function decrement from their daily maximum 5-minute exposure estimates based on the E-R function (see section 4.6.2). The extent to which differing magnitudes of exposure concentrations contribute to the total risk estimates in each year is shown in Table 5-6 for the children with asthma in Fall River and Indianapolis study areas and days with a SO_2-related increase in sRaw of 100% or more. The majority (83-100%) of the simulated individuals estimated to experience at least one day with such a lung function decrement had their 5-minute daily maximum exposure between 50 and 150 ppb. In all three study areas, there were no simulated individuals with a SO_2-related increase in sRaw of 100% or more when 5-minute daily maximum exposures were less than 40 ppb, effectively serving as a threshold for the ER function at this exposure level.

Table 5-4. Percent and number of children and adults with asthma estimated to experience at least one day per year with a SO_2-related increase in sRaw of 100% or more while breathing at an elevated rate, air quality adjusted to just meet the existing standard.

Study area	Population group	Increase in sRaw (%)	Percent (and number) of population with asthma having at least one day per year with specified increase in sRaw			
			Year 1	Year 2	Year 3	Average
Fall River	children	100	1.4 (52)	0.8 (28)	0.5 (20)	0.9 (33)
		200	0.2 (9)	0.1 (5)	<0.1 [a] (2)	0.1 (5)
	adults	100	0.3 (42)	0.2 (21)	<0.1 (9)	0.2 (24)
		200	<0.1 (6)	<0.1 (1)	0 [a]	<0.1 (2)
Indianapolis	children	100	1.5 (161)	1.3 (140)	1.1 (121)	1.3 (141)
		200	0.4 (39)	0.3 (35)	0.3 (29)	0.3 (34)
	adults	100	0.4 (158)	0.4 (147)	0.4 (128)	0.4 (144)
		200	0.1 (36)	<0.1 (32)	<0.1 (26)	<0.1 (31)
Tulsa	children	100	0	<0.1 (1)	<0.1 (1)	<0.1 (1)
		200	0	0	0	0
	adults	100	<0.1 (1)	<0.1 (2)	<0.1 (1)	<0.1 (1)
		200	0	0	0	0

[a] < 0.1 represents nonzero estimates below 0.1%. A zero (0) indicates there were no individuals having the increase in sRaw.

Table 5-5. Percent of children and adults with asthma estimated to experience multiple days per year with a SO_2-related increase in sRaw of 100% or more while breathing at elevated rate, air quality adjusted to just meet the existing standard.

Lung function decrement (increase in sRaw)	Percent (and number) of population with asthma having multiple days per year with specified increase in sRaw [a] Average per year (minimum/year – maximum/year)								
	Fall River, MA			Indianapolis, IN			Tulsa, OK		
	# Days			# Days			# Days		
	≥2	≥4	≥6	≥2	≥4	≥6	≥2	≥4	≥6
	Children, aged 5 to 18 years								
≥ 100%	0.4 (<0.1[b] - 0.7)	0.2 (<0.1-0.4)	0.1 (0 - 0.2)	0.7 (0.6 – 0.8)	0.4 (0.4)	0.3 (0.3)	no individuals experiencing multiple days with this size increase in sRaw		
≥ 200%	<0.1 (0 - 0.1)	0 [b]	0	0.2 (0.1 – 0.2)	<0.1 (<0.1)	<0.1 (<0.1)			
	Adults, aged 19 to 95 years								
≥ 100%	<0.1 (0 - <0.1)	0	0	0.2 (0.1 – 0.2)	<0.1 (<0.1)	<0.1 (<0.1)	no individuals experiencing multiple days with this size increase in sRaw		
≥ 200%	no individuals experiencing multiple days with this size increase in sRaw			<0.1 (<0.1)	<0.1 (<0.1)	0			

[a] These estimates are summarized from the single year data provided in Appendix J.
[b] < 0.1 represents nonzero estimates below 0.1%. Zero (0) indicates there were no individuals having the specified increase in sRaw.

Table 5-6. Contribution of different magnitudes of 5-minute SO_2 exposures to lung function risk (sRaw increase of at least 100%) estimated for children with asthma in Fall River.

5-minute SO_2 exposure concentration bins	Percent contribution of exposure concentration to total risk estimate [a] Fall River			Indianapolis		
	Year 1	Year 2	Year 3	Year 1	Year 2	Year 3
0 to <50 ppb	0.0%	0.0%	5.0%	0.6%	0.7%	0.8%
50 to <100 ppb	32.7%	67.9%	45.0%	37.9%	47.1%	51.2%
100 to <150 ppb	55.8%	32.1%	50.0%	45.3%	39.3%	38.8%
150 to <200 ppb	11.5%	0.0%	0.0%	2.5%	12.9%	1.7%
>=200 ppb	0.0%	0.0%	0.0%	13.7%	0.0%	7.4%

[a] These results are generated from the same data used to estimate the percent of children experiencing at least one day with an increase in sRaw ≥ 100% provided in Table 5-4.

5.4 STUDY AREA DIFFERENCES AND POPULATION DISTRIBUTION

To gain a better understanding of the role of two study area characteristics in differences of exposure estimates among the three study areas, we derived a metric that combines the census tract population counts (Figures 5-1 to 5-3) with the spatial distribution of the modeled air quality receptor design values for the air quality scenario assessed (Figures 3-6 to 3-8).[10] By merging the two variables that most influence population-based exposures – ambient concentrations and number of people – this metric can be used to indicate spatial variability in exposures and be useful in broadly comparing relative differences across the three study areas. The first section below summarizes how the exposure metric (labeled DV&POP) was derived and the subsequent section describes use of the metric in comparing the study areas.

5.4.1 Derivation of DV&POP Metric

First, the air quality receptors used for estimating exposures were identified using the APEX sites file, a file that contains the IDs for both the air quality receptors and the population census blocks used in each exposure simulation. This set of IDs was then used to link the air quality receptor design values with the census tract population data (i.e., the first 11 characters of the block IDs are the tract IDs). Second, all design values were normalized to the maximum design value in each study area, creating a set of data ranging in value of 0 to 1, with a value of 1 given to the receptor that had an original design value of 75 ppb. Because there can be multiple census blocks (and air quality receptors) within each census tract, two new ambient concentration variables were calculated using this normalized data set - the arithmetic average of all normalized receptor design values falling within each tract (nDV_{avg}) and the maximum normalized design value within each tract (nDV_{max}). This was done to represent the overall relative concentration in each tract while also recognizing the importance of the upper percentile concentrations. Third, two new population variables were created. In each study area, the tract populations were normalized by its own maximum tract population to generate values for the first population variable (and thus having a value ranging from 0 to 1 for tracts in each study area). The purpose of this population variable ($nPOP_{intra}$) was to discern intra-study area spatial differences in population. The second population variable was created similarly, though the tract populations in each study area were normalized using the maximum population in any of the three study areas ($nPOP_{inter}$). The Indianapolis study area had the tract with the greatest population, thus for this study area, the values for this second population variable were identical to values for the first population variable. However, in the other two study areas, the second

[10] Note that the degree of spatial heterogeneity of SO_2 concentrations across each study area is a function of the emission source(s) characteristics (e.g. emission rate, stack height), meteorological conditions (e.g., wind speed), among other factors.

population variable had its values relative to that of Indianapolis, thus accounting for inter-study area differences in population.

These four variables were then combined to represent the final metric (DV&POP) in each tract and weighted such that its range of values for this new metric extends from 0 to 1, as follows in equation 5-1:

$$DV\&POP = 0.333 \times nDV_{avg} + 0.333 \times nDV_{max} + 0.222 \times nPOP_{intra} + 0.112 \times nPOP_{inter}$$

Equation 5-1

We note that the weighting scheme gives more weight to the design value information (2/3) than the population data (1/3) given the significance of high concentrations for this exposure assessment. More weight was given to the intra-population variable than to the inter-population variable to allow for between study area comparisons, but focus more on within study area population variability.

5.4.2 Comparing the Study Areas with the DV&POP Metric

Census tract values of the DV&POP metric are presented in Figures 5-5 to 5-7 for the three study areas. A scale of 0 to 1 is used, varying by equivalent increments of 0.2. The highest values indicate census tracts where the confluence of population and ambient concentrations is greatest, the lowest values indicate tracts having little influence from either the population and ambient concentrations. Therefore, of greatest interest are DV&POP values at the upper end of the scale.

As a general observation, it can be seen that DV&POP values are similar in the Fall River and Indianapolis study areas and the values for those two areas differ from Tulsa (Figures 5-5 through 5-7). For example, all of the census tracts in the Fall River and Indianapolis study areas (Figures 5-5 and 5-6) have a DV&POP value greater than 0.4, while 89% of the tracts in Tulsa have DV&POP values less than 0.4, indicating that most tracts in the Fall River and Indianapolis study areas have relatively higher design values and/or populations than Tulsa census tracts. This overall observation using the metric reflects the study area-specific design value and population information for the three areas. For example, in Fall River, over 70% of receptors had hourly design values between 31 to 45 ppb (Figure 3-6) and 77% of tracts have a population greater than 3,000 people (Figure 5-1).[11] In Tulsa, by comparison, 83% of receptor sites have hourly design values less than 30 ppb (Figure 3-8) and 59% of tracts have fewer than 3,000 people (Figure 5-3).

[11] Similarly, in the Indianapolis study area, over 63% of receptors had moderately high hourly design values between 31 to 45 ppb (Figure 3-7) and 58% of tracts have a population greater than 3,000 people (Figure 5-2).

Most importantly to the exposure/risk results are tracts having the higher DV&POP values (e.g., greater than 0.6), as these are most likely locations within the study area where the highest exposures occur for the greatest number of simulated people. In Fall River, a total of 13 tracts (comprising about 73,000 people) have DV&POP values greater than 0.6, with one of these encompassing the primary source and having the highest DV&POP value of 0.82 (Figure 5-5). Similarly, in the Indianapolis study area (Figure 5-6), there are a total of 10 tracts (comprising about 86,000 people) with DV&POP values above 0.6, with the one that encompasses one of the largest sources (IPL-Harding) having the highest DV&POP value of 0.88. In contrast, the Tulsa study area (Figure 5-7) has only one tract with a DV&POP value above 0.6, and its value is 0.61.

These broad spatial differences in population size and where that might overlap with higher DVs likely contribute to the greater number of exposures at or above the benchmark levels in the Fall River and Indianapolis study areas compared with the Tulsa study area. Additionally, both Fall River and Indianapolis had a greater spatial extent of air quality receptor sites with 5-minute ambient air concentrations at or above 100 ppb than the Tulsa study area (Tables 3-14 to 3-16).

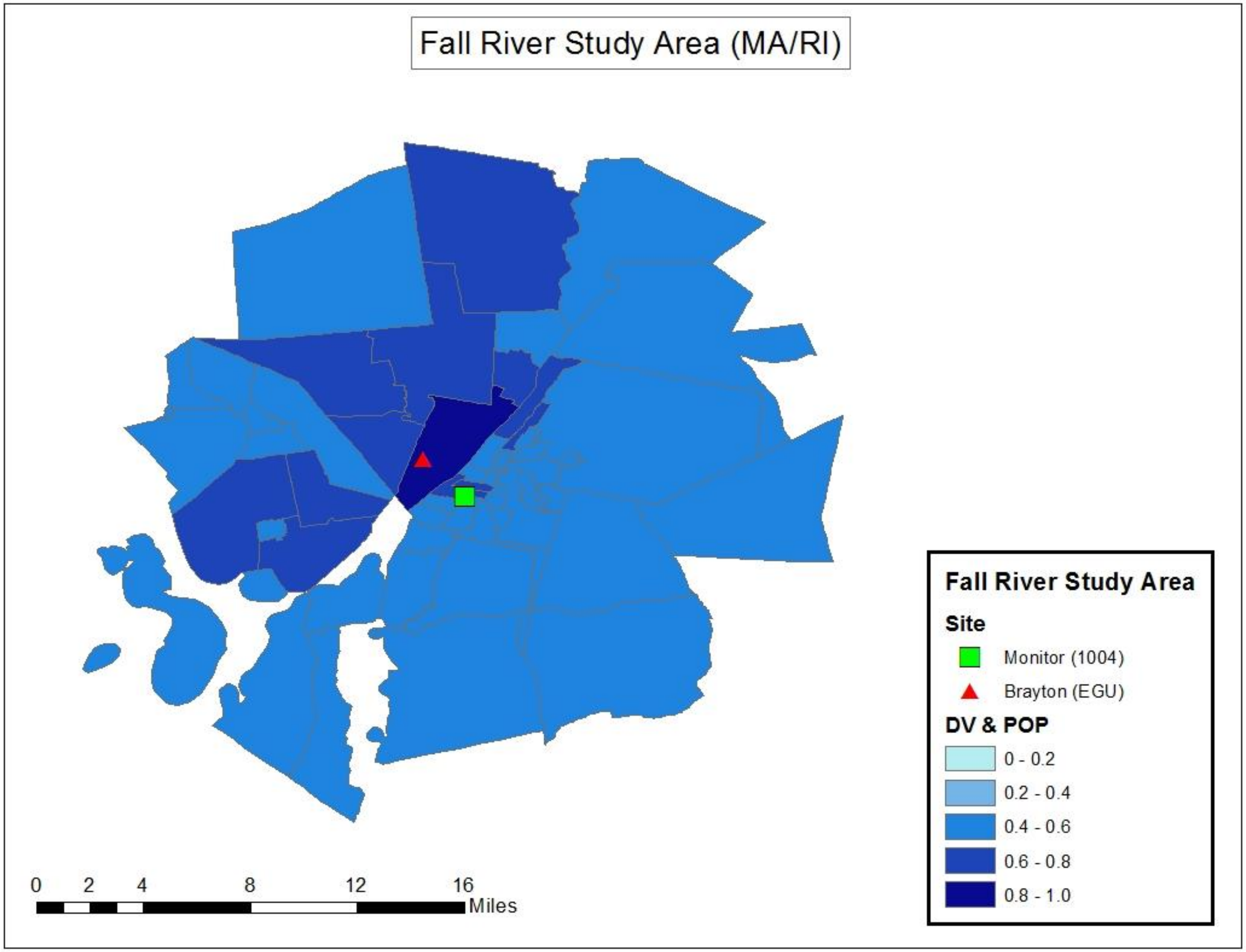


Figure 5-5. Values of the DV&POP exposure metric in the Fall River study area.

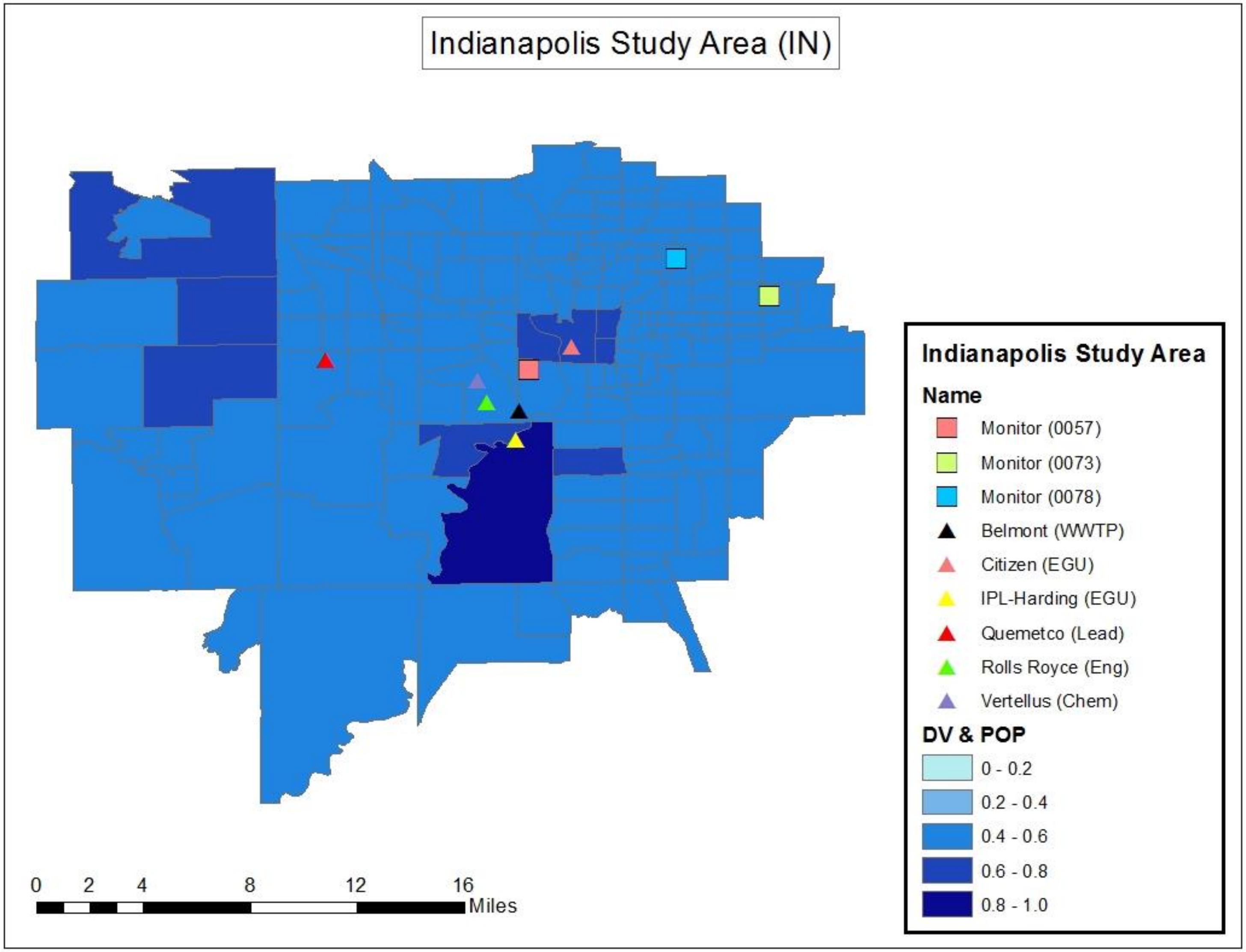


Figure 5-6. Values of the DV&POP exposure metric in the Indianapolis study area.

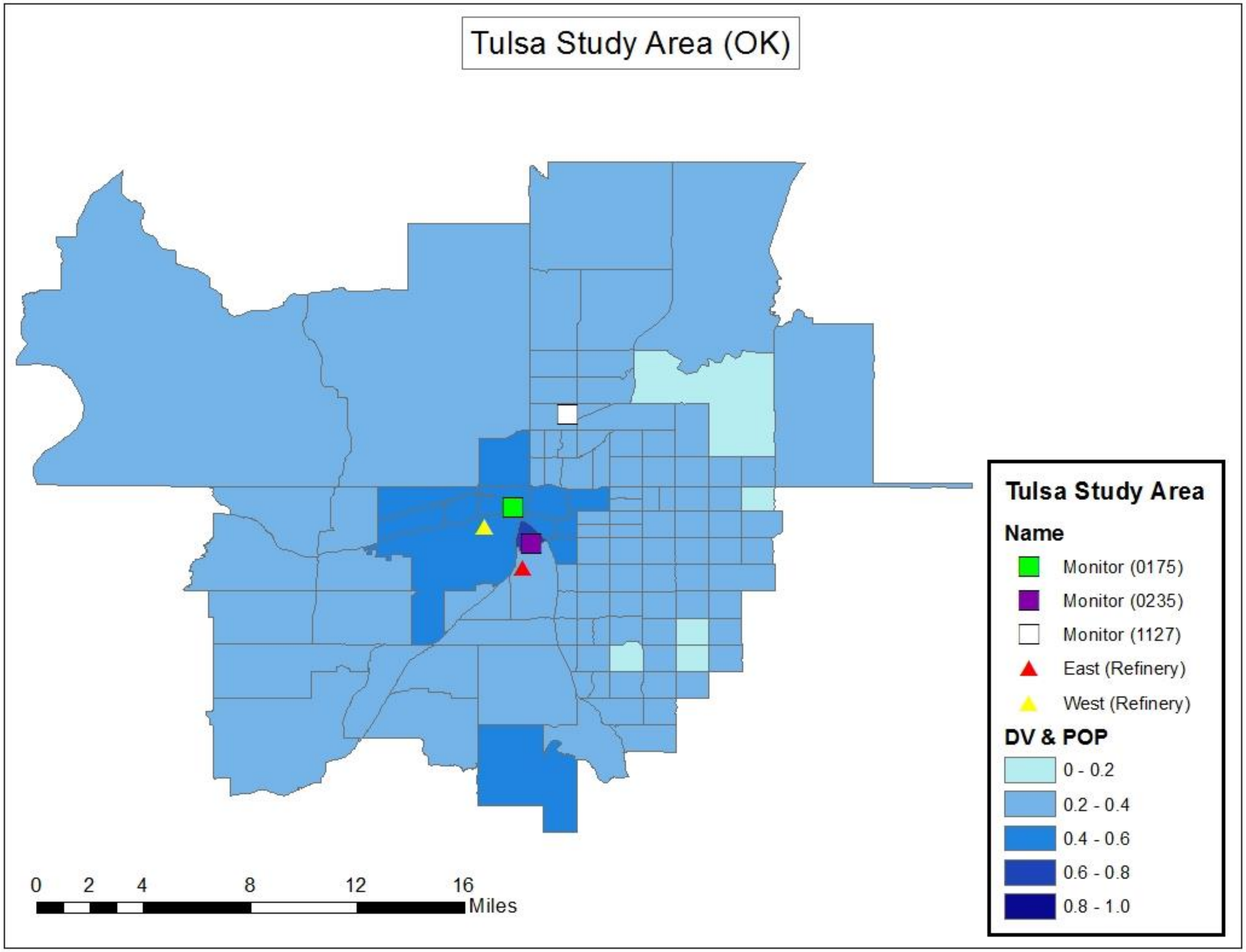


Figure 5-7. Values of the DV&POP exposure metric in the Tulsa study area.

5.5 COMPARISON WITH 2009 REA RESULTS

The results presented in this chapter and discussed above provide estimates for air quality conditions associated with just meeting the now-current 1-hour standard of 75 ppb (evaluated as 3-year average of annual 99th percentiles), an air quality scenario that was not included in the 2009 REA. As summarized in section 1.2 above, the 2009 REA included single-year air quality scenarios for 99th percentile levels of 50 ppb and 100 ppb in two study areas (St. Louis and Greene County, Missouri). For each air quality scenario, the exposure estimates for these two areas differed, and it is plausible that population and spatial heterogeneity explain those observed differences, although the type of analysis of these factors discussed in section 5.4 above was not done in the 2009 REA. Further, while the range of the exposures at or above benchmark levels estimated here is roughly consistent with the range of estimates in the 2009 REA study areas for the air quality scenarios bracketing the current standard, there are complications associated with a direct comparison of these results given the many ways in which these analyses differ from those available in the last review. In addition to the expansion in the number, type, and geographic regions of study areas assessed, there have been many improvements to input data and modeling approaches used in this assessment compared to the prior assessment, including the availability of continuous 5-minute air monitoring data at monitors within each of the three study areas. The air quality scenario in the current REA extends the time period of exposure simulations by covering a 3-year period, consistent with the statistical form established for the now-current standard. The current air quality scenario additionally focuses on the existing standard level of 75 ppb. Further, there are also differences between the current REA and the 2009 REA with regard to the air quality adjustment approach, and the methods for estimating 5-minute concentrations. Also, the years simulated in this assessment reflect more recent emissions and circumstances subsequent to adoption of the standard in 2010.

As described in section 2.2, these REA analyses are intended to be informative to EPA's consideration of potential exposures and risks that may be associated with the air quality conditions occurring under the current SO_2 standard. This is reflected in the attributes of the study areas, including the criteria used in their selection (section 3.1), the identification of specific source emissions and characteristics, local meteorological conditions, and distribution of at-risk populations. The presence in the U.S. of these areas and others having similar attributes make the findings reported here important in considering the protection provided by the SO_2 standard, as discussed in the PA.

6 VARIABILITY ANALYSIS AND UNCERTAINTY CHARACTERIZATION

An important issue associated with any population exposure or risk assessment is the characterization of variability and uncertainty. Variability refers to the inherent heterogeneity in a population or variable of interest (e.g., residential air exchange rates). The degree of variability cannot be reduced through further research, only better characterized with additional measurement. Uncertainty refers to the lack of knowledge regarding the values of model input variables (i.e., parameter uncertainty), the physical systems or relationships used (i.e., use of input variables to estimate exposure or risk or model uncertainty), and in specifying the scenario that is consistent with purpose of the assessment (i.e., scenario uncertainty). Uncertainty is, ideally, reduced to the maximum extent possible through improved measurement of key parameters and iterative model refinement.

This chapter focuses on the general characteristics of the assessment performed, including the data and approaches used to evaluate exposures and risk associated with air quality conditions that just meet the existing standard in the three study areas. The approaches used to assess variability and to characterize uncertainty in this REA are discussed in the following two sections. The primary purpose of this characterization is to provide a summary of variability and uncertainty evaluations conducted to date regarding our SO_2 exposure assessments and APEX exposure modeling and to identify the most important elements of uncertainty in need of further characterization. Each section contains a concise tabular summary of the identified components and how, for elements of uncertainty, each source may affect the estimated exposures.

6.1 TREATMENT OF VARIABILITY AND CO-VARIABILITY

The purpose for addressing variability in this REA is to ensure that the estimates of exposure and risk reflect the variability of ambient SO_2 concentrations, population characteristics, associated SO_2 exposure, and potential health risk across the study area and for the simulated at-risk populations. In this REA, there are numerous algorithms that account for variability of input data when generating the exposures or risk estimates of interest. For example, variability may arise from differences in the population residing within census blocks (e.g., age distribution) and the activities that may influence population exposure to SO_2 (e.g., time spent outdoors, performing moderate exertion-level activities outdoors). A complete range of potential exposure levels and associated risk estimates can be generated when appropriately addressing variability in exposure and risk assessments; note however that the range of values obtained would be within the constraints of the input parameters, algorithms, or modeling system used, not necessarily the complete range of the true exposure or risk values.

Where possible, we identified and incorporated the observed variability in input data sets rather than employing standard default assumptions and/or using point estimates to describe model inputs. The details regarding many of the variability distributions used in data inputs are described in Chapter 4, while details regarding the variability addressed within its algorithms and processes are found in the APEX User Guides (U.S. EPA, 2017a, b).

Briefly, APEX has been designed to account for variability in most of the input data, including the physiological variables that are important inputs to determining exertion levels and associated ventilation rates. APEX simulates individuals and then calculates SO_2 exposures for each of these simulated individuals. The simulated individuals are selected to represent a random sample from a defined population. The collection of individuals represents the variability of the target population, and accounts for several types of variability, including demographic, physiological, and human behavior. In this assessment, APEX simulated 100,000 individuals (70,000 adults and 30,000 children) to reasonably capture the variability expected in the population exposure distribution for each study area. APEX incorporates stochastic processes representing the natural variability of personal profile characteristics, activity patterns, and microenvironment parameters. In this way, APEX is able to represent much of the variability in the exposure estimates resulting from the variability of the factors effecting human exposure.

We note also that correlations and non-linear relationships between variables input to the model can result in the model producing incorrect results if the inherent relationships between these variables are not preserved. That is why APEX is also designed to account for co-variability, or linear and nonlinear correlation among several of the model inputs, provided that enough is known about these relationships to specify them. This is accomplished by providing inputs that enable the correlation to be modeled explicitly within APEX. For example, there is a non-linear relationship between the outdoor temperature and air exchange rate in homes. One factor that contributes to this non-linear relationship is that windows tend to be closed more often when temperatures are at either low or high extremes than when temperatures are moderate. This relationship is explicitly modeled in APEX by specifying different probability distributions of air exchange rates for different ambient temperatures. In any event, APEX models variability and co-variability in two ways:

- **Stochastically**. The user provides APEX with probability distributions characterizing the variability of many input parameters. These are treated stochastically in the model and the estimated exposure distributions reflect this variability. For example, the rate of SO_2 removal in houses can depend on a number of factors which we are not able to explicitly model at this time, due to a lack of data. However, we can specify a distribution of removal rates that reflects observed variations in SO_2 decay. APEX randomly samples from this distribution to obtain values that are used in the mass balance model. Further, co-variability can be modeled stochastically through the use of conditional distributions. If two or more parameters are related, conditional distributions that depend on the values of the related parameters are input to APEX. For example, the distribution of air exchange rates (AERs) in a house depends on the outdoor temperature and whether or not air conditioning (A/C) is in use. In this case, a set of AER distributions is provided to APEX for different ranges of temperatures and A/C use, and the selection of the distribution in APEX is driven by the temperature and A/C status at that time.

- **Explicitly**. For some variables used in modeling exposure, APEX models variability and co-variability explicitly and not stochastically. For example, the complete series of 5-minute ambient air SO_2 concentrations for each hour and hourly temperatures are used in model calculations. These are input to the model continuously in the time period modeled at different spatial locations, and in this way the variability and co-variability of 5-minute concentrations and hourly temperatures are modeled explicitly.

Important sources of the variability and co-variability accounted for by APEX and used for this exposure analysis are summarized in Table 6-1 and Table 6-2 below, respectively.

Table 6-1. Summary of how variability was incorporated into the exposure and risk assessment.

Component	Variability Source	Summary
Ambient Input	Meteorological data	Spatial: local surface and upper air NWS stations used. Temporal: 1-hour NWS wind data for 2011-2013, supplemented with 1-minute ASOS wind data (Appendix A).
	Emission source types and profiles	Important SO_2 emission sources include EGUs and petroleum refineries. Hourly emission profiles derived from CEMS data, where available or using EPA's 2011v6.3 emissions modeling platform combined with the SMOKE modeling system (Appendix B).
	AERMOD modeled 1-hour ambient SO_2 concentrations	Spatial: ambient SO_2 predicted to 1,400 – 1,900 air quality receptors in three geographically representative study areas Temporal: hourly SO_2 for each of three years (2011-2013).
	Ambient air monitor 5-minute concentrations	Spatial: local ambient air monitors used. Where multiple monitors available, receptors used 5-minute patterns from the closest monitor. Temporal: patterns of 5-minute continuous SO_2 concentrations within each hour used to estimate 5-minute continuous SO_2 concentrations at modeled air quality receptors.
Simulated Individuals	Population data	Individuals are randomly sampled from U.S. census blocks used in each model study area, stratified by age (single years) and sex probability distributions (U.S. Census Bureau, 2012).
	Employment	Work status is randomly generated from U.S. census data at the tract level by age and sex (U.S. Census Bureau, 2012).
	Activity pattern data	Data diaries used to represent locations visited and activities performed by simulated individuals are randomly selected from CHAD master (>55,000 diaries) using six diary pools stratified by two day-types (weekday, weekend) and three temperature ranges (< 55.0 °F, between 55.0 and 83.9 °F, and ≥84.0 °F). CHAD diaries capture real locations that people visit and the activities they perform, ranging from 1 minute to 1 hour in duration (U.S. EPA, 2017c).
	Commuting data	Employed individuals are probabilistically assigned ambient air concentrations originating from either their home or work block based on U.S. Census derived tract-level commuter data (U.S. DOT, 2012; U.S. Census Bureau, 2012).
	Longitudinal profiles	A sequence of diaries is linked together for each individual that preserves both the inter- and intra-personal variability in human activities (Glen et al., 2008).
	Asthma prevalence	Asthma prevalence is stratified by sex, single age years for children (5-17), seven adult age groups, (18-24, 25-34, 35-44, 45-54, 55-64, 65-74, and, ≥75), three regions (Midwest, Northeast, and South), and U.S. Census tract level poverty ratios (Appendix E).
Physiological Factors Relevant to Ventilation Rate	Resting metabolic rate	Five age-group and two sex-specific regression equations, use body mass and age as independent variables (Appendix H).
	Metabolic equivalents by activity (METS)	Randomly sampled from distributions developed for specific activities (some age-specific) (U.S. EPA, 2017c).
	Oxygen uptake per unit of energy expended	Randomly sampled from a uniform distribution to convert energy expenditure to oxygen consumption (U.S. EPA, 2017a, b).

Component	Variability Source	Summary
	Body mass	Randomly selected from population-weighted lognormal distributions with age- and sex-specific geometric mean (GM) and geometric standard deviation (GSD) derived from the National Health and Nutrition Examination Survey (NHANES) for the years 2009-2014 (Appendix G).
	Body surface area	Sex-specific exponential equations using body mass as an independent variable (Burmaster, 1998).
	Height	Randomly sampled from population-weighted normal distributions stratified by single age years and two sexes developed from 2009-2014 NHANES data (Appendix G).
	Ventilation rate	Event-level activity-specific regression equation using oxygen consumption rate (VO_2) and maximum VO_2 as independent variables, and accounting for intra and interpersonal variability (Appendix H).
	Fatigue and EPOC	APEX approximates the onset of fatigue, controlling for unrealistic or excessive exercise events in an individual's activity time-series while also estimating excess post-exercise oxygen consumption (EPOC) that may occur following vigorous exertion activities using several equations and input variable distributions (Isaacs et al., 2007; U.S. EPA, 2017a, b).
Microenvironmental Approach	Microenvironments: General	Five total microenvironments are represented, including those expected to be associated with high exposure concentrations (i.e., outdoors and outdoor near-road). Where this type of variability is incorporated within particular microenvironmental algorithm inputs, this results in differential exposure estimates for each individual (and event) as persons spend varying time frequency within each microenvironment and ambient air concentrations vary spatially within and between study areas.
	Microenvironments: Spatial Variability	Ambient air concentrations used in microenvironmental algorithms vary spatially within and among study areas.
	Microenvironments: Temporal Variability	All exposure calculations are performed at the event-level when using either factors or mass balance approach (durations can be as short as one minute). For the indoor microenvironments, using a mass balance model accounts for SO_2 concentrations occurring during a previous hour (and of ambient origin) to calculate a current event's indoor SO_2 concentrations.
	Air exchange rates	Several lognormal distributions are sampled based on five daily mean temperature ranges, study area region (Chapter 4) and study-area specific A/C prevalence rates from AHS survey data (U.S. Census Bureau, 2013).
	Removal rates	Values randomly selected for microenvironment-specific distributions, stratified by air conditioning usage (Chapter 4).
	Penetration factors	Indoor/outdoor ratios randomly sampled from a uniform distribution (Chapter 4).
Exposure Response Function	Regression estimates	A central tendency, along with upper and lower confidence intervals were derived using a probit function to generate a range of risk estimates.
	Exposure bins	Fine-scale bins (10-50 ppb) stratifying the population exposures were linked to the continuous E-R function.

Table 6-2. Important components of co-variability in exposure modeling.

Type of Co-variability	Modeled by APEX?	Treatment in APEX / Comments
Within-person correlations [a]	Yes	Sequence of activities performed, microenvironments visited, and general physiological parameters (body mass, height, ventilation rates).
Between-person correlations	No	Perhaps not important, assuming the same likelihood of the population of individuals either avoiding or experiencing an exposure event based on a social (group) activity.
Correlations between profile variables and microenvironment parameters	Yes	Profiles are assigned microenvironment parameters.
Correlations between demographic variables and activities	Yes	Census block demographic variables, appropriately weighted and stratified by age and sex, are used in activity diary selection.
Correlations between activities and microenvironment parameters	No	Perhaps important, but do not have data. For example, frequency of opening windows when cooking or smoking tobacco products.
Correlations among microenvironment parameters in the same microenvironment	Yes	Modeled with joint conditional variables.
Correlations between demographic variables and air quality	Yes	Modeled with the spatially varying census block demographic variables (age and sex) and fine-scale (100 m to 2 km) air quality input to APEX.
Correlations between meteorological variables and activities	Yes	Temperature is used in activity diary selection.
Correlations between meteorological variables and microenvironment parameters	Yes	The distributions of microenvironment parameters can be functions of temperature.
Correlations between drive times in CHAD and commute distances traveled	Yes	CHAD diary selection is weighted by commute times for employed persons during weekdays.
Consistency of occupation/school microenvironmental time and time spent commuting/busing for individuals from one working/school day to the next.	No	Simulated individuals are assigned activity diaries longitudinally without regard to occupation or school schedule (note though, longitudinal variable used to develop annual profile is time spent outdoors).

[a] The term correlation is used to represent linear and nonlinear relationships.

6.2 CHARACTERIZATION OF UNCERTAINTY

While it may be possible to capture a range of exposure or risk estimates by accounting for variability inherent to influential factors, the true exposure or risk for any given individual within a study area is unknown. To characterize health risks, exposure and risk assessors commonly use an iterative process of gathering data, developing models, and estimating exposures and risks, given the goals of the assessment, scale of the assessment performed, and limitations of the input data available. However, uncertainty remains and emphasis is then placed on characterizing the nature and potential magnitude of that uncertainty and its impact on exposure and risk estimates. A summary of the overall characterization is provided in section 6.2.1. The summary is followed by exposure model sensitivity analyses in section 6.2.2 that provide additional support to the characterization of four elements of uncertainty: (1) the proportional approach applied to the primary emission source to adjust ambient air concentrations to just meet the current standard, (2) estimating continuous 5-minute concentrations at ambient air monitors, (3) estimating 5-minute concentrations at modeled air quality receptors, and (4) estimating exposure and risk estimated using upper and lower bounds of the E-R function.

6.2.1 Characterizing Sources of Uncertainty

The REAs for the previous O_3, NO_2, SO_2, and CO NAAQS reviews each presented a characterization of uncertainty of exposure modeling (Langstaff, 2007; U.S. EPA, 2008, 2009a, 2010, 2014). The qualitative approach used in this and other REAs, also informed by quantitative sensitivity analyses, is described by WHO (2008). Briefly, we identified the key aspects of the assessment approach that may contribute to uncertainty in the exposure and risk estimates and provided the rationale for their inclusion. Then, we characterized the *magnitude* and *direction* of the influence on the assessment results for each of these identified sources of uncertainty.

Consistent with the WHO (2008) guidance, we scaled the overall impact of the uncertainty by considering the degree of uncertainty as implied by the relationship between the source of uncertainty and the exposure concentrations. A qualitative characterization of *low*, *moderate*, and *high* was assigned to the magnitude of influence and knowledge base uncertainty descriptors, using quantitative observations relating to understanding the uncertainty, where possible. Where the magnitude of uncertainty was rated *low*, it was judged that large changes within the source of uncertainty would have only a small effect on the assessment results. A designation of *moderate* implies that a change within the source of uncertainty would likely have a moderate (or proportional) effect on the results. A characterization of *high* implies that a small change in the source would have a large effect on results. We also included the direction of influence, indicating how the source of uncertainty was judged to potentially affect the exposure/risk estimates; this included whether the estimates were likely over-estimated ("*over*") or under-estimated ("*under*") or the direction was *unknown*. A summary of the key findings of those prior uncertainty characterizations that are most relevant to the current SO_2 exposure assessment are also provided in Table 6-3 (i.e., Langstaff, 2007; U.S. EPA, 2008, 2009a, 2010, 2014).

Table 6-3. Characterization of Key Uncertainties in Exposure and Risk Assessments using APEX.

Sources of Uncertainty		Uncertainty Characterization				Sensitivity Analysis Performed?
		Influence of Uncertainty on Exposure | Risk Estimates		Knowledge-base Uncertainty	Comments	
Category	Element	Direction	Magnitude			
Aspects of Assessment Design	Representation of SO_2 emission source types having substantial emissions	Unknown	Low - Moderate	Moderate	The three study areas include the most prevalent source type (i.e., EGUs) emitting at least 15,000 tons of SO_2 per year (95% of all U.S. facilities emitting SO_2 in 2011; U.S. EPA, 2015). There are only three other source types having emissions at least as large: copper/lead smelters (2 facilities), pulp and paper mills (2 facilities), and chemical plants (1 facility) (U.S. EPA, 2015). The limited occurrence of these large non-EGU facilities and their occurrence in locations with small populations and/or for which ambient air monitoring data for SO_2 are not available hampered their selection as study areas evaluated in this REA. To the extent that the temporal patterns of emissions and/or emissions characteristics for these source types differ in a way that would lead to greater variability in ambient SO_2 concentrations than that associated with EGU emissions, it is possible the risk/exposure estimates associated with these particular sources (if having substantial emissions) could vary from those estimated in this REA. However, risk and exposure estimates for areas with such sources would likely have limited applicability nationally due to limited prevalence of such areas across the U.S.	No
	Representation of population subgroups with asthma	Unknown	Low - Moderate	Moderate	Consistent with the ISA identification of people with asthma (and children with asthma in particular) as an important at-risk population for SO_2 in ambient air, risk estimates are developed for people with asthma and are reported separately for children and adults. Exposure and risk were not estimated for more targeted population groups with asthma based on additional personal attributes associated with increased asthma prevalence (e.g., obesity or African American or Hispanic ethnicity) generally due to limitations in the data needed to simulate such subgroups. Such data limitations affect our ability to characterize SO_2 exposure and associated health risks for different population subgroups of children and adults with asthma, some of which may have higher exposure/risk and others lower.	No

Sources of Uncertainty		Uncertainty Characterization				Sensitivity Analysis Performed?
		Influence of Uncertainty on Exposure | Risk Estimates		Knowledge-base Uncertainty	Comments	
Category	Element	Direction	Magnitude			
AERMOD Inputs and Algorithms	Algorithms (section 3.2)	Unknown	Low	Low	Multiple historical model evaluations consistently demonstrate unbiased ambient air concentrations under variety of conditions. Some potential dispersion scenarios may not be adequately represented and in some instances, concentration variability could be under-represented by model algorithms, however, it is largely unknown as to how this uncertainty might apply in this application.	No
	Meteorological Data (section 3.2.1.1 and Appendix A)	Unknown	Low – Moderate	Low	A limited number of missing hours of wind data remain in dataset, potentially leading to under-estimation. Model predictions have low to medium sensitivity to surface roughness characteristics, as long as they are appropriate for the site of the meteorological data inputs. Data are from a well-known and quality-assured source. One minute ASOS wind data used to supplement 1-hour data for improved completeness, reducing the number of calms and missing data. Two meteorological stations (one upper and one surface) are used to represent meteorological conditions in each study area, some of which are located a few to several km from ambient air monitor sites and the modeled air quality receptors. There is uncertainty in the extent to which conditions measured at these stations represent study area meteorological conditions, particularly wind speed and direction, and how this could affect the estimation of hourly and 5-minute concentration variability.	No
	Stationary Source Emissions and Profiles (section 3.2.2 and Appendix B)	Both	Low	Low	Temporal emission characteristics are well represented for most modeled point sources. Most temporal data are from a well-known quality-assured source of direct measurements.	No
Ambient Air Monitor Concentrations	Database Quality	Both	Low	Low	All ambient pollutant measurements available from AQS are comprehensive and subject to quality control. Completeness criteria applied to hourly concentrations ensure air quality representativeness.	No

Sources of Uncertainty		Uncertainty Characterization				Sensitivity Analysis Performed?
		Influence of Uncertainty on Exposure | Risk Estimates		Knowledge-base Uncertainty	Comments	
Category	Element	Direction	Magnitude			
	Missing Data Substitution (section 3.5.1)	Under	Low	Low	Missing ambient air concentration values (hourly, 5-minute maximum, 5-minute continuous) were interpolated using a statistical technique. Use of this type of approach is appropriate for data sets having a limited missing number of total values (<5-10%), though will constrain substituted values within the bounds of the measured concentrations. In addition, there are a few monitors missing concentrations for several hours/minutes per day and for several days in the year, most notably in the Indianapolis study area (Table 3-9), potentially missing a few high concentration events (if actually occurred) that would not be estimated using the interpolation technique. However, completeness of the maximum 5-minute concentrations was reasonable (<10%) for all monitors used in estimating 5-minute concentrations, thus the missing within hour 5-minute concentrations are of lesser importance and likely contribute less uncertainty.	No
	Estimation of Continuous 5-minute Concentrations (section 3.5.2)	Under	Low	Low	For one year in Fall River (2013), only the 5-minute maximum measurements within each hour were reported. A series of lognormal distributions were used to estimate the 5-minute continuous patterns occurring with each hour for these monitors (Section 3.5.2). Excellent agreement was observed comparing the estimated versus the measured values for each of the hourly and 5-minute maximum concentrations. Agreement between the estimated and measured 5-minute continuous concentrations was also excellent, though exhibiting some deviations (Figure 3-6). In addition, the estimated 5-minute continuous concentrations had less overall variability compared to the measurement data (Table 3-10). However, there was negligible difference in exposures when comparing an APEX simulation that used measured continuous 5-minute concentrations versus one that used estimated values.	Yes, section 6.2.2.1
	Temporal Representation (section 3.5.2 and 3.5.3)	Both	Low	Low	Temporal scale (5-minutes) is appropriate for analysis performed. Monitored hourly and 5-minute maximum data are screened for temporal completeness and considered appropriate. While 5-minute continuous data were not screened for completeness, the number of missing values were limited in most study areas and for most years (Table 3-9).	No

Sources of Uncertainty		Uncertainty Characterization				Sensitivity Analysis Performed?
		Influence of Uncertainty on Exposure | Risk Estimates		Knowledge-base Uncertainty	Comments	
Category	Element	Direction	Magnitude			
	Spatial Representation (section 3.5.3.1)	Both	Moderate	Moderate	There were few ambient air monitors available to approximate 5-minute patterns across study area: Fall River, one monitor available and used to estimate 5-minute concentrations; Indianapolis, three monitors available (two were used); Tulsa, four monitors available (three were used). Where more than one monitor was available, the air quality receptors used 5-minute concentration patterns from closest monitor.	No
Air Quality Receptor Concentrations	Concentration Used to Represent Sources Not Modeled (section 3.2.4)	Both	Moderate	Low - Moderate	There is uncertainty in the estimates of hourly concentration associated with SO_2 emission sources not explicitly modeled in the three study areas. While temporal variability in these estimates is accounted for by calculating diurnal and seasonal values, year to year variability is not considered, thus not accurately accounting for instances where the contribution may vary by year. The value used for each hour/season is the 3-year average of the 99th percentile concentration (section 3.2.4), an approach that at most times would generally tend to overestimate these concentrations. Further, monitor hours that may have concentrations influenced by modeled sources were identified for exclusion using wind direction data from nearby airports. This provided a consistent approach across study areas as local wind direction data were not reported at all monitors. However, uncertainty is contributed in circumstances where the airport wind direction does not reflect conditions occurring at a monitor. The magnitude of such uncertainty may be sizeable at some monitors.	No

Sources of Uncertainty		Uncertainty Characterization				Sensitivity Analysis Performed?
		Influence of Uncertainty on Exposure | Risk Estimates		Knowledge-base Uncertainty	Comments	
Category	Element	Direction	Magnitude			
	AERMOD Predicted Hourly Concentrations	Both	Low - Moderate	Moderate	Overall, comparisons of model predicted hourly values and ambient air measurements in each study area indicate good agreement when considering concentration magnitude alone. The first of the three years tended to exhibit the highest concentrations in addition to having the best agreement across all three study areas. In the Fall River study area, AERMOD over-predicted low to mid percentile concentrations and under-predicted upper percentile concentrations. In the Indianapolis study area, AERMOD over-predicted all percentiles of the concentration distribution at the monitor closest to the primary emissions source, while under-predicting mid to high percentile concentrations at the monitors more distant from the primary emissions source. In the Tulsa study area, AERMOD under-predicted mid to high percentile concentrations at the monitor closest to the primary emissions source, while over-predicting most percentile concentrations at monitors more distant from the primary emissions source (Appendix D). Such differences are of lesser importance to the assessment estimates given the focus on air quality after adjustment to just meet the existing standard.	No
	Hourly Ambient Air Concentration Estimates during Times of Relatively Greater Exposure Potential	Both	Low - Moderate	Low - Moderate	As separately concluded for the generalized performance evaluation summarized in Appendix D, these comparisons that consider both spatial and temporal variability (i.e., where and when peak concentrations occur) also indicate reasonable agreement between the model estimates and measurements at the nearby monitor site(s). Similarity in the paired concentrations across much of the respective distributions for times when exposure potential may be greatest provide additional positive support for concluding the modeled air quality surfaces are likely useful for estimating exposures in this REA (section 3.2.5 and Appendix K). However, having limited monitoring data available in each study area and the inability to directly evaluate the concentrations for the air quality scenario that is the focus of this REA limits the extent by which conclusions can be made regarding model performance in estimating spatial variability in hourly concentrations for that scenario.	No

Sources of Uncertainty		Uncertainty Characterization				Sensitivity Analysis Performed?
		Influence of Uncertainty on Exposure | Risk Estimates		Knowledge-base Uncertainty	Comments	
Category	Element	Direction	Magnitude			
	Adjustment of Hourly Concentrations to Just Meet the Existing Standard: Proportional Approach for Primary Source (section 3.4)	Under	Low	Moderate	Performance of this approach in this REA depends in part on the degree of proportionality in the air quality distribution and the magnitude of the ambient air concentration adjustment. A proportional approach was judged adequate for such a use (section 3.4 above; REA Planning Document; Rizzo, 2008). The approach used in this REA is a modification of 2009 REA adjustment approach in that the proportional adjustment was applied only to the concentration contribution from the primary emission source in each study area, holding concentrations contributed from all other sources as is. The sharpness of the concentration gradient from the primary emission source relative to the other emission sources could be an important factor in determining the impact to the adjusted air quality surface. However, sensitivity analyses that modified the air quality receptor having the maximum design value (section 6.2.2.2) indicate there was negligible (in two areas) or somewhat limited (in the third area) impact to the estimated exposures by varying the magnitude of the adjustment and the number of receptors to which the adjustment was applied.	Yes, section 6.2.2.2
	Approach Used to Estimate 5-minute Concentrations: Linking 5-minute Monitor to Hourly Receptor Concentrations (section 3.5.3)	Both	Low - Moderate	Moderate	Hourly concentrations modeled at the air quality receptors were linked to the 5-minute monitor concentrations using the rank order of the hourly concentrations. Two alternative approaches were developed and evaluated. The first, a calendar based approach, linked the modeled receptor concentrations to the monitor by date and hour of day. The second used hourly concentration bins (i.e., 5 ppb increments). There were differences when comparing the upper percentiles of the 5-minute concentration distributions, particularly when comparing the calendar based approach to the rank order and binning approaches. There were also notable differences to the percent of the at-risk population exposed at or above benchmarks when comparing results from the three adjustment approaches. However, little difference was observed when comparing risk of lung function decrements estimated using each of these three approaches.	Yes, section 6.2.2.3

Sources of Uncertainty		Uncertainty Characterization				Sensitivity Analysis Performed?
		Influence of Uncertainty on Exposure | Risk Estimates		Knowledge-base Uncertainty	Comments	
Category	Element	Direction	Magnitude			
	Estimated Peak 5-minute Concentrations during Times of Relatively Greater Exposure Potential	Both	Low-Moderate	Moderate	As concluded in section 3.5.3.3, there is reasonable agreement between the 5-minute concentrations estimated at air quality receptors and measurement concentrations at the local ambient air monitor site(s), considering concentrations of interest and the number of days at or above these concentrations. Having limited monitoring data available in each study area however limits our ability to assess the reasonableness of the degree to which these concentrations may be present spatially. It appears that the spatial extent of receptors having the highest 5-minute concentrations could be over-estimated in the Indianapolis study area and less so (to possibly not at all) in the Fall River and Tulsa study areas.	No
APEX: General Input Databases	Population Demographics and Commuting (sections 4.1.1 and 4.3.2)	Both	Low	Low	Comprehensive and subject to quality control. Differences in 2010 population data versus modeled years (2011-2013) are likely small when estimating percent of population exposed.	No
	Activity Patterns (CHAD) (sections 4.3.1 and 4.3.3)	Both	Low - Moderate	Low - Moderate	Comprehensive and subject to quality control. Increased number of diaries used to estimate exposure from 2009 SO_2 REA. Thoroughly evaluated trends and patterns in historical activity pattern data – no major issues noted with use of historical data to represent current patterns (Figures 5G-1 and 5G-2 of U.S. EPA, 2014). Compared outdoor event participation and outdoor time of CHAD diary data with larger American Time Use Survey (ATUS) data – CHAD participation is higher than ATUS, likely due to ATUS survey methods. Comparison of activity data (outdoor events and exertion level) for people with asthma generally similar to individuals without asthma (Table 4-5) (see also Tables 5G2-to 5G-5 of U.S. EPA, 2014). There is little indication of differences in time spent outdoors comparing activity patterns across U.S. regions, though sample size may be a limiting factor in drawing significant conclusions (U.S. EPA, 2014). Remaining uncertainty exists for other influential factors that cannot be accounted for (e.g., SES, region/local participation in outdoor events and associated amount of time).	No

Sources of Uncertainty		Uncertainty Characterization				Sensitivity Analysis Performed?
		Influence of Uncertainty on Exposure | Risk Estimates		Knowledge-base Uncertainty	Comments	
Category	Element	Direction	Magnitude			
	Meteorological (NWS) (section 4.2)	Both	Low	Low	Comprehensive and subject to quality control, having very few missing values. Limited use in selecting CHAD diaries for simulated individuals and AERs that may vary with temperature. However, while using three years of varying meteorological conditions, the 2011-2013 MET data set may not reflect the full suite of conditions that could exist in future hypothetical air quality scenarios or across periods greater than 3-years.	No
	Asthma Prevalence Weighted by Poverty Status (section 4.1.2 and Appendix E)	Both	Low	Low - Moderate	Data used are from peer-reviewed quality controlled sources. Use of these data accounts for variability in important influential variables (poverty status, as well as age, sex, and region). It is possible that variability in microscale prevalence is not entirely represented when considering other potentially influential variables such as race and obesity, two attributes that can influence asthma prevalence and can vary spatially (section 4.1.2). Family income level was used in this REA to represent spatial variability in asthma prevalence and may, in some instances, capture spatial variability in race and obesity (Ogden et al., 2010), and thus to some extent, reasonably represent the potential influence race and obesity have on asthma prevalence. However, instances where these influential variables are not fully represented in simulating the at-risk population, and where populations identified by such variables are associated with increased asthma prevalence that may spatially intersect with the highest ambient concentrations, could lead to uncertainty in estimated exposures and health risk. Further characterization could be appropriate by comparing with local prevalence rates stratified by a similar collection of influential variables, where such data exist.	No
APEX: Microenvironmental Concentrations	Vehicle PE Factors (Section 4.4.5)	Both	Low	Moderate	Input distribution is from an older measurement study and for a different pollutant (section 4.4.5). Considering that the exposures of interest need to be concomitant with elevated exertion, the accurate estimation of 5-minute exposures occurring inside vehicles is considered relatively unimportant.	No

Sources of Uncertainty		Uncertainty Characterization				Sensitivity Analysis Performed?
		Influence of Uncertainty on Exposure | Risk Estimates		Knowledge-base Uncertainty	Comments	
Category	Element	Direction	Magnitude			
	Indoor: Air Exchange Rates (sections 4.4.1 and 4.4.3)	Both	Low	Moderate	Uncertainty due to random sampling variation via bootstrap distribution analysis indicated the AER geometric mean (GM) and standard deviation (GSD) uncertainty for a given study area tends to range from ±1.0 GM and ± 0.5 GSD hr^{-1} (Langstaff, 2007). Each of the three study areas (Fall River, Indianapolis, and Tulsa) used AER from a geographically similar city (New York, Detroit/New York, and Houston, respectively). Non-representativeness remains an important issue as city-to-city variability can be wide ranging (GM/GSD pairs can vary by factors of 2-3) and data available for city-specific evaluation are limited (Langstaff, 2007). There is uncertainty associated with the use of an AER derived from a different city than the REA study areas. That said, indoor microenvironments are considered less likely to contribute to an individual's daily maximum 5-minute SO_2 exposure while at elevated exertion levels and likely does not contribute substantially to uncertainty in the exposure and risk estimates.	No
	Indoor: A/C Prevalence (section 4.4.2)	Both	Low	Low	Data were obtained from a reliable source, are comprehensive, and subject to quality control (US Census Bureau, 2013). For two of the three study areas (Fall River and Indianapolis), data from a geographically related city were used (Boston and Louisville, respectively). There is uncertainty associated with the use of an AC prevalence derived from a different city than the REA study areas. That said, indoor microenvironments are considered less likely to contribute to an individual's daily maximum 5-minute SO_2 exposure while at elevated exertion levels and likely does not contribute substantially to uncertainty in the exposure and risk estimates.	No
	Indoor: Removal Rate (section 4.4)	Unknown	Low	Moderate	In the 2009 REA it was found that indoor exposures may be underestimated when not using all 5-minute concentrations within the hour, an issue resolved in this current REA by using estimates of all 5-minute values. Data used to develop removal rates were obtained from a comprehensive review, though many assumptions were needed in developing the distributions. However, most peak exposures concomitant with elevated exertion are expected to occur outdoors, thus accurate estimation of indoor concentrations is of reduced importance.	No

Sources of Uncertainty		Uncertainty Characterization				Sensitivity Analysis Performed?
		Influence of Uncertainty on Exposure | Risk Estimates		Knowledge-base Uncertainty	Comments	
Category	Element	Direction	Magnitude			
APEX: Simulated Activity Profiles	Longitudinal Profiles (section 4.3.4)	Under	Low - Moderate	Moderate	The magnitude of potential influence for this uncertainty would be mostly directed toward estimates of multiday exposures. Simulations indicate the number of single day and multiday exposures of interest can vary based on the longitudinal approach selected (Che et al., 2014). As discussed in chapter 4, the D&A method provides a reasonable balance of this exposure feature. Note however, long-term diary profiles (i.e., monthly, annual) do not exist for a population, thus limiting the evaluation. Further, the general population-based modeling approach used for main body REA results does not assign rigid schedules, for example explicitly representing a 5-day work week for employed people.	No
	Commuting (section 4.3.2)	Both	Low	Moderate	Method used in this assessment is designed to link Census commute distances with CHAD vehicle drive times. Considered an improvement over the prior approach that did not match commute distance and activity time. While vehicle time is accounted for through diary selection, it is not rigidly scheduled. However, accurate estimation of exposures occurring while inside vehicles is considered unimportant because it is unlikely to occur at elevated exertion.	No
	Activity Patterns for At-Risk Population (section 4.3.3)	Both	Low	Low - Moderate	Analyses of activity patterns of people with asthma are similar to that of individuals not having asthma (section 4.3.3; see also Tables 5G-2 to 5G-5 of U.S. EPA, 2014).	No
APEX: Physiological Processes	Body Weight (NHANES) (section 4.1.3.1 and Appendix G)	Unknown	Low	Low	Comprehensive and subject to quality control, appropriate years (2009-2014) selected for simulated population, though possible small regional variation is not represented by national data.	No

Sources of Uncertainty		Uncertainty Characterization				Sensitivity Analysis Performed?
		Influence of Uncertainty on Exposure | Risk Estimates		Knowledge-base Uncertainty	Comments	
Category	Element	Direction	Magnitude			
	RMR (section 4.1.3.2, Appendix H)	Unknown	Low	Low	New, improved algorithm used for this assessment. Comprehensive literature review resulted in construction of large data base used to derive algorithm. Algorithm considers variables most influential to RMR (i.e., age, body weight, and sex). There are other factors that could affect intra-personal variability in RMR such as time-of-day (Haugen et al., 2003) or seasonal/temperature influences (van Ooijen et al., 2004; Leonard et al., 2014). Variability from these and other potentially influential factors may be indirectly accounted for by the residual error term used in the RMR Equation 4-2 depending on the extent to which these influential factors varied across the clinical study data that were used to create the RMR analytical data set. However, because there is inadequate information regarding the presence of multiple RMR measurements for individual study subjects, we could not estimate intra-personal variability nor could we use these influential factors, other than age and sex, as explanatory variables in the RMR equation. Therefore, any influences on spatial variability in RMR, both within and among the three study areas, would largely be driven by the spatial distribution of age and sex.	No
	METS Distributions (section 4.1.3.2)	Over	Low - Moderate	Moderate	APEX estimated daily mean METs range from about 0.1 to 0.2 units (between about 5-10%) higher than independent literature reported values (Table 15 of Langstaff, 2007). However, shorter-term values are of greater importance in this assessment, thus METs could be better characterized where short-term METS data are available.	No
	Ventilation Rates (section 4.1.3.3 and Appendix H)	Unknown	Low	Low - Moderate	Predictions made using the prior algorithm showed excellent agreement with independent measurement data, particularly when considering simulated study group (Graham and McCurdy, 2005; Figure 5-23 and Figure 5-24 of U.S. EPA, 2014). New algorithm derived using the same data observed to have improved predictability (Appendix H). However, a shorter-term comparison (5-minutes or a single hour rather than daily) of predicted versus measured ventilation rates, while more informative, cannot be performed due to lack of ventilation rate data at this duration and considering influential factors (e.g., age, particular activity performed).	No

Sources of Uncertainty		Uncertainty Characterization				Sensitivity Analysis Performed?
		Influence of Uncertainty on Exposure | Risk Estimates		Knowledge-base Uncertainty	Comments	
Category	Element	Direction	Magnitude			
	EVR Characterization of Moderate or Greater Exertion (section 4.1.3.3)	Both	Moderate	Moderate	Given that the EVR serves as a cut point for selecting individuals performing moderate or greater exertion activities and is an approximated mean value applied to the population as a whole, the simulated number of people achieving this level of exercise (along with having a response) could be either under or overestimated. This is because EVR is calculated as a function of body surface area as a means to extrapolate the ventilation rates achieved by adults in the controlled human exposure studies to that of our simulated children. Fundamentally, the EVR assumes that differences between adults and children with regard to the target ventilation rate denoting ventilation for moderate or greater exertion can be described by differences in body surface area (and hence body weight). On average, body surface area for adults is approximately 1.95 m^2 and for children is approximately 1.33 m^2 (U.S. EPA, 2011), indicating that on average, the ventilation rate required to meet an EVR of 22 (i.e., moderate or greater exertion) is about 43 L/min and 29 L/min for adults and children, respectively. Recommended ventilation rates representing moderate intensity exercise (3<METS<6) in children is about 88% that of adults (U.S. EPA, 2011) and is close to that estimated using the above VE or body surface area differences (68%), Based on this, simulated children use a lower ventilation threshold to reach moderate or greater exertion relative to that of adults, possibly leading to overestimation of health risks, holding all other potential influential factors constant. Additionally, it is possible that there is a distribution of EVRs within a particular age (or age group) that vary based on body weight or perhaps even fitness level (e.g., individuals having the same body weight but different fat free mass). For example, among people with asthma, an appreciable portion are obese. Obesity is an important personal attribute that can affect activity patterns and EVR, both influential variables in estimating SO_2 risk. Based on the role of body weight in energy expenditure alone (e.g., without consideration of the role inter-individual variation in fitness level has on VE for different activities), simulated obese people (BMI≥30) would need to have higher VE than non-obese people performing the same activity to achieve an EVR of 22 given their relatively greater body surface area (on average, approximately 30-60% greater). APEX estimates ventilation rates for obese individuals that are approximately 10-30% greater than non-obese individuals largely the result of obese individuals having, on average, a 10-40% greater resting metabolic rate than non-obese individuals. Accordingly, SO_2-related health risk estimates derived for simulated obese people with asthma could be underestimated, holding all other potential influential factors constant.	No

Sources of Uncertainty		Uncertainty Characterization				Sensitivity Analysis Performed?
		Influence of Uncertainty on Exposure | Risk Estimates		Knowledge-base Uncertainty	Comments	
Category	Element	Direction	Magnitude			
Lung Function Risk Estimation (section 4.6.2)	Risk Estimation for Exposures Below 100-200 ppb	Over	Low - Moderate	Low - Moderate	While there is very strong support for SO_2 being causally linked to lung function responses within the range of tested exposure levels (i.e., ≥ 200 ppb), data are limited or lacking for lower concentrations. Data available at 100 ppb are limited to studies in which SO_2 was administered by mouthpiece, some of which also do not include a control exposure to clean air while exercising (Sheppard et al. 1981; Sheppard et al., 1984; Koenig et al., 1989; Koenig et al., 1990; Trenga et al., 2001). These studies indicate smaller responses (in adults and adolescents) than is observed in the 200 ppb chamber exposures. No data are available at lower exposure levels below 100 ppb. Since this assessment assumes there is a causal relationship at levels below 100 ppb, the influence of this source of uncertainty would be to over-estimate risk.	No
	Probit Model Used to Estimate E-R Function	Unknown	Low	Low	It was necessary to estimate responses at SO_2 levels both within the range of exposure levels tested (i.e., 200 to 1,000 ppb) as well as below the lowest exposure levels used in free-breathing controlled human exposure studies (i.e., below 200 ppb). We have developed probabilistic exposure-response relationships using a probit form, considered appropriate for this assessment. However, the regression model assumes a positive response occurring at any exposure concentration, of particular relevance to the lowest exposures.	No
	Use of E-R data from Studies of Individuals having Mild/Moderate Asthma to Represent Any Asthma Severity	Unknown	Unknown	Moderate	The data set that was used to estimate exposure-response relationships included people with mild and/or moderate asthma. There is uncertainty with regard to how well the population of people with mild and moderate asthma included in the series of SO_2 controlled human exposure studies represent responses that might be expected across the entire distribution of people with asthma in the U.S. population. As indicated in the ISA (section 5.2.1.2), the subjects studied do not generally include people with asthma that would be classified as severe by today's classification standards. The available studies "suggest that adults with moderate/severe asthma may have more limited reserve to deal with an insult compared with individuals with mild asthma" (ISA, p. 5-22).	No

Sources of Uncertainty		Uncertainty Characterization				Sensitivity Analysis Performed?
		Influence of Uncertainty on Exposure | Risk Estimates		Knowledge-base Uncertainty	Comments	
Category	Element	Direction	Magnitude			
	Reproducibility of SO_2-induced Lung Function Response	Unknown	Unknown	Low	The risk assessment assumes that the SO_2-induced responses for individuals are reproducible. We note that this assumption has some support in that one study (Linn et al., 1987) exposed the same subjects on two occasions to 0.6 ppm and the authors reported a high degree of correlation ($r > 0.7$ for people with mild asthma and $r > 0.8$ for people with moderate asthma, $p < 0.001$), while observing much lower and nonsignificant correlations ($r = 0.0 – 0.4$) for the lung function response observed in the clean air with exercise exposures.	No
	Use of E-R Derived from Adults for Children	Unknown	Unknown	Low - Moderate	Because the vast majority of controlled human exposure studies investigating lung function responses were conducted with adult subjects, the risk assessment relies on data from adult subjects with asthma to estimate exposure-response relationships that have been applied to all individuals with asthma, including children aged 5-18. The available evidence includes some studies of adolescents (aged 12-18) with asthma that indicate generally similar effects as observed for adults, although precise comparisons are not feasible with the available data (ISA, pp. 5-22 to 5-23). The studies involving adolescents administered SO_2 via inhalation through a mouthpiece rather than an exposure chamber. This technique bypasses nasal absorption of SO_2 and can result in an increase in lung SO_2 uptake. Given this is a limited dataset and the lack of any such studies for children younger than 12, , the uncertainty in the risk estimates for children with asthma is greater than those for adults.	No
	SO_2 Exposure History	Both	Low	Moderate	The risk assessment assumes that the SO_2-induced response on any given day is independent of previous SO_2 exposures and only the highest daily 5-minute exposure (under moderate or greater exertion) is assessed. The limited evidence related to this source indicates effects from a subsequent-day exposure to not be statistically significantly different from the first day. Further, responses to repeated exposures within an hour have been found to be diminished responses from initial ones, although data are limited or lacking regarding exposures repeated after multiple hours but within the same 24-hour period (ISA, section 5.2.1.2).	No

Sources of Uncertainty		Uncertainty Characterization				Sensitivity Analysis Performed?
		Influence of Uncertainty on Exposure | Risk Estimates		Knowledge-base Uncertainty	Comments	
Category	Element	Direction	Magnitude			
	Assumed No Interaction of other Co-pollutants on SO_2-related Lung Function Responses	Under	Low	Moderate	There are a few studies regarding the potential for an increased response to SO_2 when exposure is in the presence of other common pollutants such as PM (potentially including particulate sulfur compounds), nitrogen dioxide and ozone, although the studies are limited (e.g., with regard to relevance to ambient exposure concentrations) and/or provide inconsistent results (ISA, p. 5-25; 2008 ISA, section 3.1.4.7; ISA, pp. 5-143 to 5-144). For example, "studies of mixtures of particles and sulfur oxides indicate some enhanced effects on lung function parameters, airway responsiveness, and host defense," however, "some of these studies lack appropriate controls and others involve [sulfur-containing species] that may not be representative of ambient exposures" (ISA, p.5-144).	No

6.2.2 Exposure Model Sensitivity Analyses

6.2.2.1 Continuous 5-minute Concentrations – Estimated versus Measured

Analyses evaluating the approach used to estimate the twelve 5-minute concentrations for each hourly concentration in the assessment is summarized in section 3.5.2. These analyses utilized datasets at monitors for which continuous 5-minute data are available; the analyses indicate reasonable agreement between the estimated and measured concentrations. By design, the estimated hourly and within-hour 5-minute maximum concentrations were identical to the measured hourly and 5-minute maximum concentrations, though sampling from lognormal distributions led to instances where the within-hour pattern of the eleven other estimated within-hour 5-minute concentrations varied from that measured (Figure 3-6).

We evaluated the impact this difference may have on exposures in the Fall River study area, the only study area that used this method to estimate continuous 5-minute concentrations for the single year that continuous 5-minute measurements were not available (2013). Two identical APEX simulations were performed in the Fall River study area that differed only by the ambient air concentrations used for input to the model. Both simulations used a single air quality district, the center of which was the location of monitor 250051004, and employed a 10 km radius of influence to select the census blocks comprising the exposure modeling domain. One simulation used the continuous 5-minute concentrations measured in 2011 at the ambient air monitor and the other using the pattern of 5-minute continuous concentrations estimated for that same year and location (and initiated by the monitor's measured hourly and daily maximum 5-minute concentrations). All other model settings were the same as that used for the APEX simulations performed for the main REA, though only children with asthma were simulated.

We first evaluated statistics of interest beyond those presented in Table 3-10. Of interest were the upper percentile concentrations and number of times the 5-minute ambient air concentrations were at or above the benchmark concentrations. Table 6-4 provides the results of this analysis. Consistent with results provided in chapter 3, there are differences between estimated and measured values at the upper percentile concentrations shown here (i.e., 99^{th} percentile of the distribution and the number of values at or above 100 ppb), with the estimated percentile concentrations slightly lower than the percentile concentrations for the measured values. However, in the APEX simulation results there is little to no difference in either the estimated exposures at or above the benchmarks (Table 6-5) or in the percent of the children expected to experience a lung function decrement (Table 6-6) when considering the varying concentration input.

Table 6-4. Comparison of measured and estimated continuous 5-minute SO_2 concentrations in ambient air, Fall River monitor 250051004, 2011.

Monitor ID 250051004		
Continuous 5-minute SO_2 concentrations (ppb)		
Percentile of distribution	**Estimated**	**Measured**
p0	0.0	0.0
p1	0.0	0.0
p5	0.1	0.1
p10	0.4	0.5
p25	1.0	1.1
P50	1.8	1.9
p75	2.8	2.7
p90	5.5	5.2
p95	9.4	9.0
p99	34.1	36.6
p100	241.1	241.1
Number of times per year 5-minute concentration at or above benchmark		
Benchmark Concentration (ppb)	**Estimated**	**Measured**
100	144	147
200	5	5
300	0	0
400	0	0

Table 6-5. Comparison of simulated exposures, for children with asthma, at or above benchmarks using measured versus estimated continuous 5-minute SO_2 concentrations from monitor 250051004, Fall River, 2011.

benchmark (ppb)	5-minute ambient air concentrations	Percent of children with asthma having exposures at or above 5-minute benchmark concentration					
		number of days per year at or above benchmark concentration					
		≥ 1	≥ 2	≥ 3	≥ 4	≥ 5	≥ 6
100	Measured	43.9	20.0	9.0	3.9	1.6	0.7
	Estimated	43.2	19.2	8.4	3.7	1.5	0.7
200	Measured	8.3	0.6	<0.1 [a]	0 [a]	0	0
	Estimated	8.3	0.6	<0.1	0	0	0
300	Measured	no individuals estimated to experience any days at or above 300 ppb					
	Estimated						

[a] < 0.1 represents nonzero estimates below 0.1%. A zero (0) indicates there were no individuals having the specified exposure.

Table 6-6. Comparison of simulated lung function decrements in children with asthma using measured versus estimated 5-minute continuous SO_2 concentrations, Fall River 2011.

sRaw	5-minute ambient air concentration input	Percent of children with asthma estimated to experience one or more days with an increase of sRaw of specified amount					
		number of days per year					
		≥ 1	≥ 2	≥ 3	≥ 4	≥ 5	≥ 6
100%	Measured	2.4	0.8	0.4	0.3	0.2	0.1
	Estimated	2.3	0.9	0.4	0.3	0.2	<0.1 [a]
200%	Measured	0.6	0.1	0 [a]	0	0	0
	Estimated	0.6	0.1	0	0	0	0

[a] < 0.1 represents nonzero estimates below 0.1%. A zero (0) indicates there were no individuals having the specified exposure.

6.2.2.2 Adjustment of Hourly Concentrations to Just Meet the Existing Standard

In this assessment, a proportional approach was used to adjust air quality to just meet the current standard. For the exposure and risk results presented in Chapter 5, as described in section 3.4, we adjusted concentrations for the source contributing the most to the air quality receptor concentrations, and that single receptor having the maximum design value in each study area. Thus, all other design values calculated for the modeled receptors in the study area following the air quality adjustment were less than 75 ppb, with one receptor having a design value of 75 ppb.

In light of the variation in adjustment factors (Table 3-8), the fact that the factor is derived from the highest design value, and the finding that, while the model predicted hourly concentrations were found generally comparable with monitor measurements, there were a few instances where the highest upper percentile concentrations could be overestimated (see Appendix D, Table D-3), we have evaluated the impact on the estimated population exposures of an alternative adjustment approach. The alternative approach is intended to address the potential for overestimation at the few highest-concentration receptors that could result in the application of an overly large adjustment factor for a number of the receptors in the modeling domain. This alternative adjustment procedure modifies the selection of the receptor that is used to calculate the adjustment factor. Rather than select the single maximum design value to determine the adjustment factor for all receptor concentrations within a study area, we chose the 99th percentile design value to determine the adjustment factor for the receptor with that design value, and for receptors with lower values. Thus, all receptors having design values less than the 99th percentile following the air quality adjustment would have a design value less than 75 ppb. All study area receptors having design values above the 99th percentile design value were adjusted using their own individual adjustment factors that resulted in each of them having adjusted concentrations that also yielded a design value of 75 ppb.

Table 6-7 summarizes the adjustment factors used in this alternative approach. The air quality scenario created by this alternative approach, just like the base approach used for the exposure and risk results in Chapter 5, reflects air quality conditions that just meet the existing standard. However, this alternative adjustment procedure using the 99th percentile design value results in a greater spatial distribution of relatively higher concentrations across the study area compared with the scenario created using the maximum design value, which leads to higher percentages of children with asthma having exposures above benchmark concentrations and lung function decrements. Figures 6-1 to 6-3 illustrate this in each of the study areas, showing the overlay of the population distribution and the design values resulting from the two different adjustment approaches.

Table 6-7. Air quality adjustment factors for main body REA and sensitivity analysis.

Study area	Approach for Main body REA		Alternative Approach for Sensitivity Analysis		
	Maximum Design value (ppb)	Factor applied to all receptors	99th percentile design value (ppb)	Factor applied to Receptors < 99th percentile design value	Factor applied to Receptors > 99th percentile design
Fall River	101.4	1.46	83.2	1.12	1.14 – 1.46
Indianapolis	311.3	4.21	205.2	2.77	2.85 – 4.21
Tulsa	73.5	0.98	63.1	0.82	0.81 - 0.98

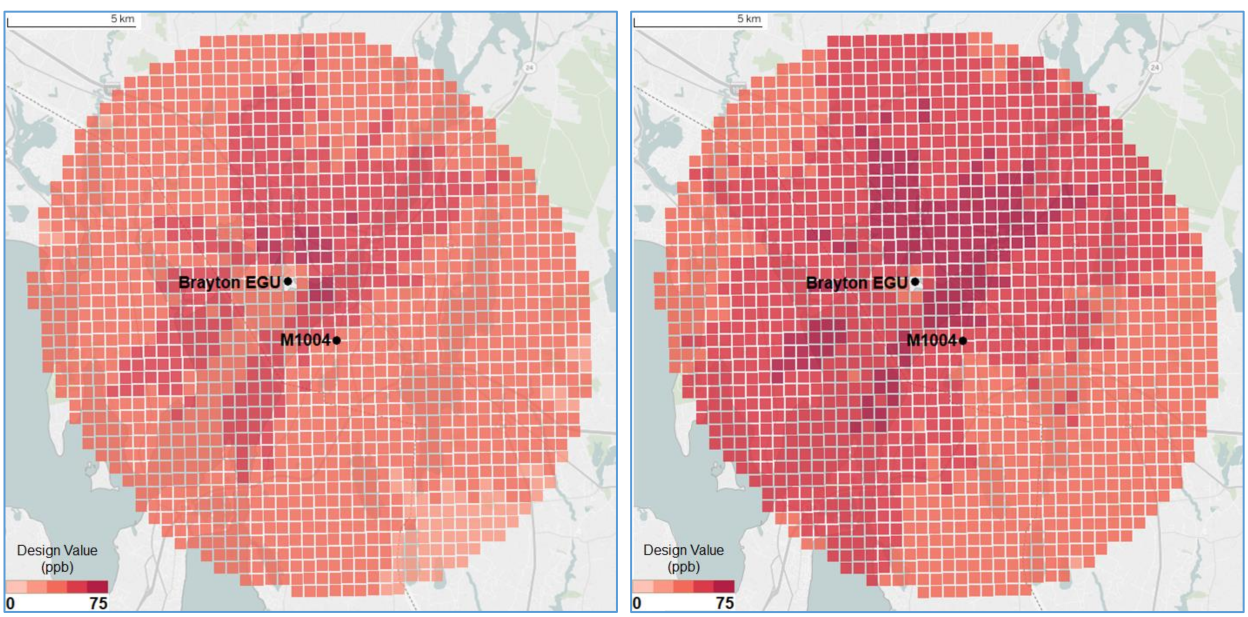


Figure 6-1. Spatial pattern of design values using an adjustment based on the maximum design value (left panel) and an adjustment based on the 99th percentile design value (right panel) in the Fall River study area.

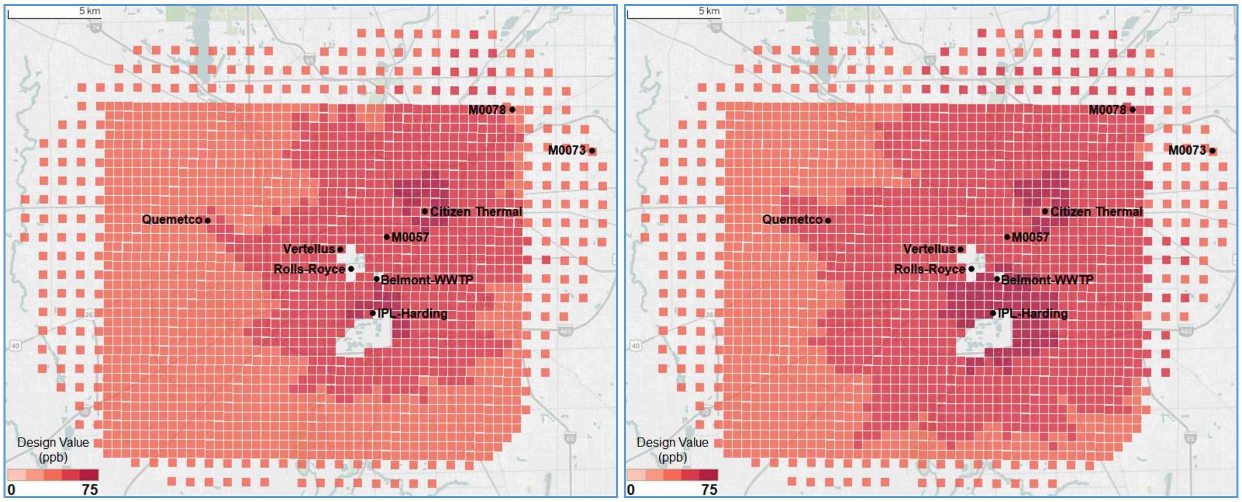


Figure 6-2. Spatial pattern of design values using an adjustment based on the maximum design value (left panel) and an adjustment based on the 99th percentile design value (right panel) in the Indianapolis study area.

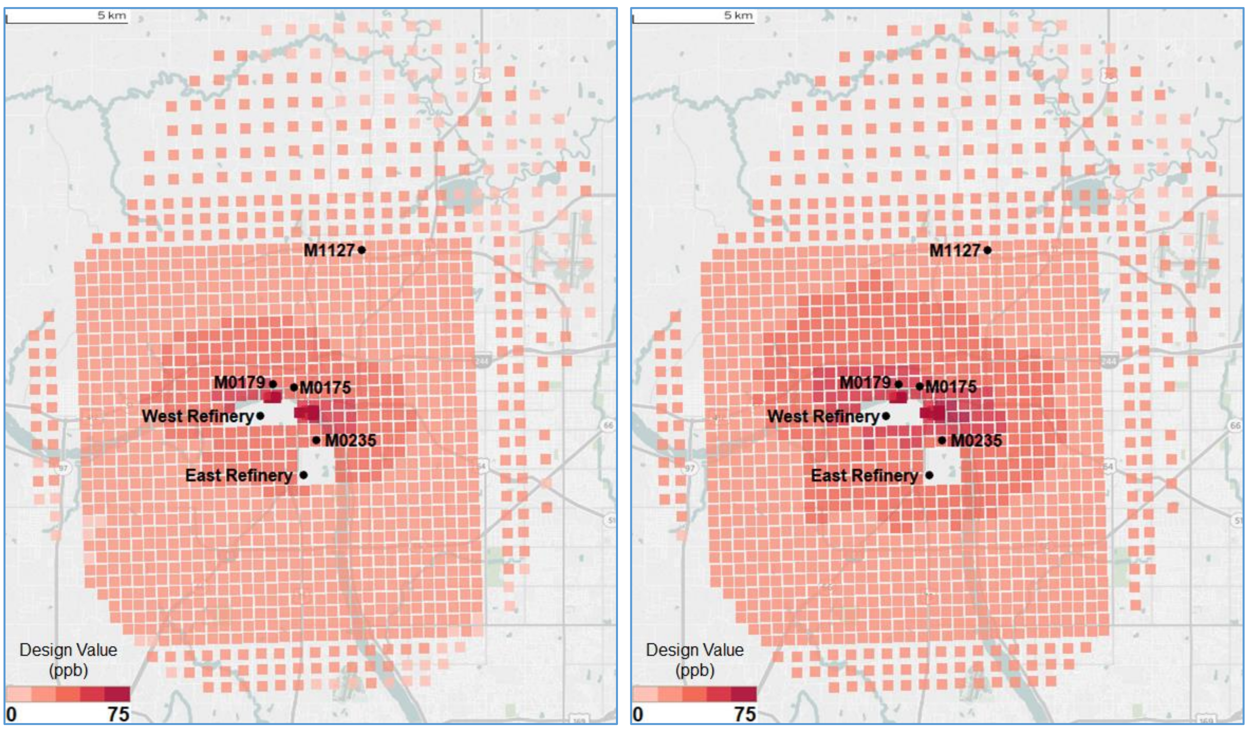


Figure 6-3. Spatial pattern of design values using an adjustment based on the maximum design value (left panel) and an adjustment based on the 99th percentile design value (right panel) in the Tulsa study area.

We performed APEX simulations using these air quality data sets derived with the alternative adjustment approach, and holding all model settings identical to those used to generate the exposures presented in Chapter 5. Exposures and risk of lung function decrements were estimated for children with asthma in the three study areas for all three years. Tables 6-8 through 6-11 present the results of these alternative simulations, including a comparative summary of the results provided in Chapter 5. In general, there is a greater percent of children expected to experience at least one daily maximum exposure at or above the benchmark concentrations when using the alternative adjustment based on the 99th percentile design value compared to that estimated using the adjustment based on the maximum design value (Table 6-8). The difference was most noticeable for the Fall River study area, particularly considering the 100 ppb benchmark (i.e., 7 to 14 percentage points at the mean and maximum, respectively). The difference was smaller when considering the 200 ppb benchmark in the Fall River study area and both benchmarks in the two other study areas (i.e., mainly fractions of a percentage point difference for any simulation). Further, there was also a greater percent of multiple exposures at or above the 100 ppb benchmark in the Fall River study area using the alternative adjustment approach, although the difference was limited to a few percentage points (Table 6-9). Only a fractional difference in the percent of children experiencing multiple exposures at or above the 100 ppb benchmark was observed for the Indianapolis study area, and there was little to no difference observed in any study area or when considering multiple exposures at or above the 200 ppb benchmark.

When considering lung function risk estimated using the two different adjusted air quality surfaces, results using the 99th percentile design value for the adjustment are similar to those estimated using the adjustment approach employing the maximum design value, although differing slightly for the Fall River study area (Table 6-10 and 6-11). On average, about 1% of children are estimated to experience at least one or multiple days with an increase in sRaw at or above 100% in the Fall River and Indianapolis study areas, regardless of the adjustment approach. Results for the Tulsa study area indicate few (<0.1%) to no children estimated to experience any increase in sRaw of interest, neither single nor multiple days. Little to no difference was observed for increases in sRaw at or above 200% in any study area when considering the alternative adjustment approach, neither single nor multiple days.

Table 6-8. Comparison of two approaches used to adjust ambient air concentrations to just meet the existing standard (2011-2013): Percent of children with asthma estimated to experience at least one day per year with a SO_2 exposure at or above 5-minute benchmark concentrations while at elevated exertion.

Study area	Benchmark Concentration (ppb) [a]	Percent of children with asthma having at least one day per year ≥ benchmark concentration: mean (min – max)	
		Max DV used to adjust air quality [b]	Max 99th DV used to adjust air quality
Fall River	100	19.4 (12.3 – 32.7)	26.7 (13.8 – 46.8)
	200	<0.1 [c] (0 [c] – 0.2)	0.7 (0 – 2.2)
Indianapolis	100	22.4 (18.0 - 27.0)	23.0 (18.8 – 26.1)
	200	0.7 (0 – 1.0)	0.6 (0 – 1.0)
	300	0.3 (0 – 0.8)	0.2 (0 – 0.7)
	400	0.1 (0 – 0.3)	<0.1 (0 – 0.2)
Tulsa	100	0.1 (<0.1 – 0.2)	0.4 (0 – 0.8)
	200	0	0

[a] There were no daily maximum 5-minute exposures at or above 300 ppb benchmark in any study area.
[b] Data from Table 5-2.
[c] < 0.1 represents nonzero estimates below 0.1%. A value of zero (0) indicates there were no individuals having the selected exposure in any year.

Table 6-9. Comparison of two approaches used to adjust ambient air concentrations to just meet the existing standard (2011-2013): Percent of children with asthma estimated to experience multiple days per year with a SO_2 exposure at or above 5-minute benchmark concentrations while at elevated exertion.

Study area	Benchmark Concentration (ppb)	Percent of children with asthma having multiple days per year > benchmark concentration: mean (min – max)					
		Max DV used to adjust air quality [a]			Max 99th DV used to adjust air quality		
		>2 days	>4 days	>6 days	>2 days	>4 days	>6 days
Fall River	100	5.5 (1.6 – 12.2)	0.9 (<0.1 [b] – 2.6)	0.2 (0 [b] – 0.6)	10.5 (2.0 – 24.0)	2.8 (0.1 – 7.7)	1.0 (0 – 2.8)
	200	no results included multiple days per year at or above this benchmark concentration			<0.1 (0 – <0.1)	0	0
Indianapolis	100	6.8 (4.7 – 8.0)	0.8 (0.3 – 1.0)	0.1 (<0.1 – 0.2)	6.9 (5.3 – 7.9)	1.0 (0.6 – 1.3)	0.2 (0.1 – 0.3)
	200	no results included multiple days per year at or above this benchmark concentration					
Tulsa	100	no results included multiple days per year at or above this benchmark concentration			<0.1 (0 - <0.1)	0	0
	200	no results included multiple days per year at or above this benchmark concentration					

[a] Data from Table 5-3.
[b] < 0.1 represents nonzero estimates below 0.1%. A value of zero (0) indicates there were no individuals having the selected exposure in any year.

Table 6-10. Percent of children with asthma estimated to experience at least one day per year with a SO_2-related increase in sRaw of 100% or more while breathing at elevated rates, air quality adjusted to just meet the existing standard.

Study area	sRaw (%)	Percent of children with asthma having at least one day per year > sRaw level: mean (min – max)	
		Max DV used to adjust air quality [a]	Max 99th DV used to adjust air quality
Fall River	100	0.9 (0.5 – 1.4)	1.1 (0.6 – 1.9)
	200	0.1 (<0.1 [b] – 0.2)	0.2 (<0.1 – 0.4)
Indianapolis	100	1.3 (1.1 – 1.5)	1.3 (1.1 – 1.5)
	200	0.3 (0.3 – 0.4)	0.3 (0.3 – 0.4)
Tulsa	100	<0.1 (0 [b] – <0.1)	<0.1 (<0.1 – <0.1)
	200	There were no individuals that experienced a day with this size increase in sRaw	

[a] Data from Table 5-4.
[b] < 0.1 represents nonzero estimates below 0.1%. A value of zero (0) indicates there were no individuals having the selected exposure in any year.

Table 6-11. Percent of children with asthma estimated to experience multiple days per year with a SO_2-related increase in sRaw of 100% or more while breathing at elevated rates, air quality adjusted to just meet the existing standard.

Study area	Lung function decrement (increase in sRaw)	Percent of children with asthma having multiple days per year ≥ sRaw level: mean (min – max)					
		Max DV used to adjust air quality [a]			Max 99th DV used to adjust air quality		
		≥2	≥4	≥6	≥2	≥4	≥6
Fall River	≥ 100%	0.4 (<0.1[b] – 0.7)	0.2 (<0.1 – 0.4)	0.1 (0[b] – 0.2)	0.6 (0.2 – 1.0)	0.2 (<0.1 – 0.4)	0.1 (<0.1 – 0.3)
	≥ 200%	<0.1 (0 – 0.1)	0	0	<0.1 (0 – 0.2)	0	0
Indianapolis	≥ 100%	0.7 (0.6 – 0.8)	0.4 (0.4)	0.3 (0.3)	0.7 (0.7 – 0.8)	0.4 (0.4 – 0.5)	0.3 (0.3)
	≥ 200%	0.2 (0.1 – 0.2)	<0.1 (<0.1)	<0.1 (<0.1)	0.2 (0.1 – 0.2)	<0.1 (<0.1)	<0.1 (<0.1)
Tulsa	There were no individuals experiencing an sRaw at or above any level of interest for multiple days						

[a] Data from Table 5-5.
[b] < 0.1 represents nonzero estimates below 0.1%. A value of zero (0) indicates there were no individuals having the selected exposure in any year.

6.2.2.3 Estimating 5-minute Concentrations at Air Quality Receptors

The approach that has been used to relate the continuous 5-minute concentrations based on ambient air measurement data to the 1-hour modeled air quality receptor concentrations is the rank order of the hourly concentration distributions (rank-order distribution approach), as summarized in section 3.5.2 (and also evaluated in section 6.2.2.1). To inform our consideration of uncertainty associated with this approach, we have also evaluated two alternative approaches: a calendar-based and concentration bin-based approach. Sensitivity analyses comparing this approach to the two alternatives that were considered are described here.

The calendar-based approach uses the actual date and time of each of the two concentration datasets (monitor and modeled) as the linking variable. Thus, the temporal patterns in hourly (and hence 5-minute patterns) would be the same at all the modeled air quality receptors, though normalized by their respective hourly concentrations that occur during that same hour (effectively employing equation 3-3, though instead of the rank order to match hourly concentrations, the consecutive calendar date and hour-of-day are used). We did not use the calendar-based approach to develop the air quality surfaces used in generating the main body exposure and risk estimates because we felt it would not appropriately represent the patterns in 5-minute concentrations, given the relationship between the within-hour 5-minute concentration variability and the magnitude of the hourly concentrations. Often times, there is greater variability in the 5-minute concentrations occurring at low hourly concentrations (particularly hourly values less than 1 ppb) than at higher hourly concentrations (U.S EPA, 2009, section 7.2.3.2). Further, we also expected that the monitor(s) would not necessarily reflect the exact temporal pattern that could occur at all receptors simultaneously, given the generally sporadic nature of peak concentrations driven by temporal and spatial variability in meteorology. That said, this mismatching of the temporal patterns observed at the monitor with the air quality receptors using the calendar-based approach would likely lead to instances where the 5-minute concentrations at the upper percentiles of the distribution are overestimated (i.e., assigning greater variability in 5-minute concentrations obtained from low hourly ambient air monitor concentrations to the highest hourly modeled concentrations). Alternatively, 5-minute concentrations at the lower percentiles would tend to be underestimated in certain instances. As the estimated risk is largely a function of the highest exposures, avoiding the potential assignment of increased variability to higher modeled concentrations was a factor in not selecting this approach for the REA. Nevertheless, how this selection affected the exposure and risk results warranted additional evaluation as given here.

The second alternative approach to assigning 5-minute variability to the hourly concentrations, the concentration bin-based approach, used the actual concentration levels in each of the two concentration datasets. Both monitor and modeled hourly concentrations were binned by 5 ppb increments, except for the lowest hourly concentrations (i.e., three bins were used – a 0 concentration bin, the second for hourly concentrations between 0 and 1 ppb, then a third for hourly concentrations between 1 and 5). This concentration bin-based approach is similar to that using the rank order distribution approach, though likely improves the matching of the hourly concentrations between the two concentration data sets, where different (i.e., structurally the monitor hourly concentration distribution becomes more like the receptor hourly concentration distribution). One limitation to the concentration bin-based approach is that there could be limits to the monitor data set in providing measurement data to all of the bins, particularly the highest hourly concentrations in the air quality scenario of interest for this REA (conditions just meeting the existing standard). This was the case in the monitoring data for the Fall River and Tulsa study areas, where the 2011-2013 monitor design values were 64 and 55 ppb, respectively. Hence, nearly all the hourly concentrations were also below the existing standard level of 75 ppb in these two study areas. Therefore, the pattern of the 5-minute concentrations associated with the highest hourly concentrations in those areas would all rely on very few measurements, leading to uncertainty in their estimation.

Table 6-12 provides the statistics calculated for the upper percentiles of the 5-minute concentrations, for air quality adjusted to just meet the existing standard, derived using each of the three methods: (1) the rank order distribution approach (used in the assessment); (2) the calendar-based approach; and (3) the concentration bin-based approach. The table presents the 5-minute concentrations estimated at all air quality receptor locations, along with statistics calculated using the ambient air monitor measurement data (also adjusted to meet the standard). Consistent with what was described above, the calendar-based approach results in unusually high 5-minute concentrations, with several receptors exhibiting concentrations at or above 300 ppb. Neither the monitor nor the receptors using the rank order distribution-based approach had concentrations at or above 300 ppb, while the concentration bin-based approach yielded a few receptors (i.e., about 15 or more) with estimated 5-minute concentrations at or above that level.

Table 6-12. Comparison of three approaches for using continuous 5-minute monitoring data to estimate 5-minute concentrations associated with modeled 1-hour concentrations at receptor locations: Air quality adjusted to just meet the existing standard, Fall River study area 2011.

Statistic	5-minute SO_2 Concentrations in Ambient Air (ppb)			
	Estimation Approach			Adjusted monitoring data at monitor location
	Calendar	Rank Order Distribution	Binned	
p90p90	12	11	11	5
p99p90	12	11	11	
maxp90	12	11	11	
p90p99	29	32	31	37
p99p99	38	41	40	
maxp99	45	48	48	
p90max	303	183	236	241
p99max	459	247	338	
maxmax	662	268	386	

Abbreviations: p90 = 90th percentile of 5-minute concentrations at monitor. p90p90 = 90th percentile of the distribution of all study area receptor 90th percentile 5-minute concentrations. Etc.

For this sensitivity analysis, all three of these approaches were used to generate an air quality surface of 5-minute concentrations in the Fall River study area and used to simulate exposures of children with asthma for 2011. All other model settings and input data were held the same as in the main analysis in Chapter 5; the only difference among these three simulations was the 5-minute concentration input described above. Table 6-13 shows the resulting estimated exposures at or above the selected benchmarks. The largest differences among the three approaches are estimates for the 100 ppb benchmark. There are greater percentages of children with asthma estimated to experience at least one day with an exposure at or above 100 ppb using the calendar-based and concentration bin-based approaches than using the rank order distribution approach. There is less variability across the three approaches when considering three or more days with exposures at or above this benchmark. Consistent with the greater number of estimated 5-minute ambient concentrations at or above the higher benchmarks (200 through 400 ppb), the calendar-based approach is the only approach estimating any days with exposures above these benchmarks. Given the discussion provided above regarding this particular approach, these results using the calendar-based approach are likely overestimates of exposure.

Table 6-13. Comparison of three approaches for using continuous 5-minute ambient air monitoring data to estimate 5-minute concentrations associated with modeled 1-hour concentrations: Estimated exposures for air quality adjusted to just meet the existing standard, Fall River, 2011.

benchmark concentration (ppb)	5-minute concentration approach	Percent of children with asthma estimated to experience one or more days with exposures at or above 5-minute benchmark concentration, while breathing at elevated rates					
		Number of days per year					
		≥1	≥2	≥3	≥4	≥5	≥6
100	Calendar-based	37.2	15.4	6.9	3.6	1.8	1.0
	Rank order distribution [a]	32.7	12.2	5.5	2.6	1.3	0.6
	Concentration bin-based	36.9	14.7	6.6	2.9	1.4	0.8
200	Calendar-based	4.7	0.5	0.1	0 [b]	0	0
	Rank order distribution	0.2	0	0	0	0	0
	Concentration bin-base	1.2	<0.1 [b]	0	0	0	0
300	Calendar	1.4	<0.1	0	0	0	0
	Rank order distribution	0	0	0	0	0	0
	Concentration bin-based	0	0	0	0	0	0
400	Calendar-based	0.3	0	0	0	0	0
	Rank order distribution	0	0	0	0	0	0
	Concentration bin-based	0	0	0	0	0	0

[a] Data from Table 5-2, Table 5-3, and Appendix J.
[b] < 0.1 represents nonzero estimates below 0.1%. A value of zero (0) indicates there were no individuals having the selected exposure in any year.

Table 6-14 shows the percent of children with asthma estimated to experience at least one or more days per year with a SO_2-related increase in sRaw of 100% or more while breathing at elevated rates, using the three different approaches. The general pattern of results is similar as for the benchmark comparison, and indicates low frequency of occurrence of lung function decrements on at least one day or multiple days (all ≤ 2%), at both levels of interest.

Table 6-14. Comparison of three approaches for using continuous 5-minute monitoring data to estimate 5-minute concentrations associated with modeled 1-hour concentrations: Estimated lung function decrements associated with exposure to air quality adjusted to just meet the existing standard, Fall River 2011.

Lung function decrement (increase in sRaw)	5-minute concentration approach	Percent of children with asthma estimated to experience one or more days with specified response [a]					
		number of days per year					
		>1	>2	>3	>4	>5	>6
100%	Calendar-based	2.0	0.7	0.4	0.2	0.2	0.1
	Rank order distribution	1.4	0.7	0.5	0.4	0.3	0.2
	Concentration bin-based	1.6	0.7	0.4	0.3	0.2	0.2
200%	Calendar-based	0.4	<0.1 [b]	0 [b]	0	0	0
	Rank order distribution	0.2	0.1	<0.1	0	0	0
	Concentration bin-based	0.3	0.1	<0.1	0	0	0

[a] Data from Table 5-4, Table 5-5, and Appendix J.
[b] < 0.1 represents nonzero estimates below 0.1%. A value of zero (0) indicates there were no individuals having the selected exposure in any year.

6.2.2.4 E-R Function for Lung Function Risk Estimates

The E-R functions for lung function risk were generated from the controlled human study data provided in Table 4-12 using a probit regression (as described in section 4.6.2 above). In addition to estimates for the risks of increases in sRaw of at least 100% and 200% based on the best fit (mean) probit model regression coefficients, we also generated lower and upper bounds for estimated risks using the 5th and 95th percentile predictions of the regression coefficients (see section 4.6.2 and Appendix J, Table J-28).

For the presentation here, the lower and upper bound E-R functions were combined with the distribution of exposures estimated in each study area, as was done using the mean regression estimates to generate the risk estimates presented in section 5.3. As for many of the sensitivity analyses in this chapter, the focus of this presentation is on risks for children with asthma experiencing 5-minute exposures while breathing at elevated rates. The estimated risks using each of the three E-R functions (for each of the two severities of response) averaged across the 3-year study period are provided in Table 6-15.

The risks estimated for the three functions vary as expected. The highest risks (both for single occurrences as well as multiple occurrences) are derived using the upper bound function and the lowest with the lower bound function. With regard to the Fall River and Indianapolis risk estimates, the differences in risk estimated using the upper bound function versus that estimated using the mean E-R function, in terms of percent of the population, are about 2 to 3 percentage points when considering the estimate of children experiencing at least one day per year with an increase in sRaw of at least 100%. The differences are smaller for multiple such occurrences (e.g., a 1.4 to 1.8 percentage point difference at most considering two or more days in a year), and also for occurrences of a 200% increase in sRaw (at most a 1.2 percentage point difference considering at least one day per year).

Table 6-15. Comparison of estimated lung function risk using mean, lower bound and upper bound of the fitted E-R function: Percent of children with asthma estimated to experience at least one or multiple days per year with a SO_2-related increase in sRaw of 100% or more while breathing at elevated rates, air quality adjusted to just meet the existing standard, 2011-2013.

Study Area	Lung function decrement (increase in sRaw)	E-R Function [a]	Percent of children with asthma estimated to experience one or more days with an increase of sRaw of specified amount (average across 3-year period)					
			Number of days per year					
			≥1	≥2	≥3	≥4	≥5	≥6
Fall River	100%	LB	0.2	<0.1 [b]	<0.1	0 [b]	0	0
		Mean [c]	0.9	0.4	0.3	0.2	0.1	0.1
		UB	2.7	1.8	1.3	1.1	0.9	0.8
	200%	LB	There were no children that experienced a day with an increase in sRaw of at least 100% using this E-R function					
		Mean	0.1	<0.1	<0.1	0	0	0
		UB	1.1	0.7	0.5	0.4	0.4	0.3
Indianapolis	100%	LB	0.5	0.2	0.1	<0.1	<0.1	<0.1
		Mean	1.3	0.7	0.5	0.4	0.4	0.3
		UB	3.5	2.5	2.0	1.8	1.6	1.4
	200%	LB	<0.1	0	0	0	0	0
		Mean	0.3	0.2	0.1	<0.1	<0.1	<0.1
		UB	1.5	1.1	0.9	0.8	0.7	0.7
Tulsa	100%	LB	There were no children that experienced a day with an increase in sRaw of at least 100% using either E-R function					
		Mean	<0.1	0	0	0	0	0
		UB	0.5	0.3	0.2	0.2	0.2	0.1
	200%	LB	There were no children that experienced a day with an increase in sRaw of at least 200% using either E-R function					
		Mean						
		UB	0.2	0.1	0.1	<0.1	<0.1	<0.1

[a] LB is a lower bound E-R function derived using the 5th percentile for the mean regression coefficient, Mean is the E-R function representing the best fit (mean) regression estimate, UB is an upper bound E-R function derived using the 95th percentile for the mean regression coefficient, each derived using the controlled human exposure-response study data in Table 4-12. See also section 4.6.2 and Figure 4-4.

[b] < 0.1 represents nonzero estimates below 0.1%. A value of zero (0) indicates there were no individuals having the selected exposure in any year.

[c] From main body REA results Tables 5-4 and 5-5 and Appendix J.

Risk estimated using the lower bound E-R function yields a smaller percent of children compared to that using the mean E-R function, with at most a 0.8 percentage point difference for children experiencing at least one day per year with an increase in sRaw of at least 100% in the Fall River and Indianapolis study areas.

Regarding the Tulsa study area, there were no children estimated to experience a 100% increase in lung function decrement when using the lower bound function; with the mean E-R functions, fewer than 0.1% were estimated to experience at least one day with an SO_2-related increase in sRaw of 100%. When using the upper bound function to estimate risk, a fraction of a percent (all ≤0.5%) of children with asthma were estimated to experience at least one or multiple days per year with a SO_2-related increase in sRaw of 100%.

REFERENCES

Burmaster DE. (1998). Lognormal distributions for skin area as a function of body weight. Risk Analysis. 18(1):27-32.

Glen G, Smith L, Isaacs K, McCurdy T, Langstaff J. (2008). A new method of longitudinal diary assembly for human exposure modeling. J Expos Sci Environ Epidem. 18:299-311.

Haugen HA, Melanson EL, Tran ZV, Kearney JT, Hill LO. (2003). Am J Clin Nutr. 78:1141-1144.

Isaacs K, Glen G, McCurdy T., and Smith L. (2007). Modeling energy expenditure and oxygen consumption in human exposure models: Accounting for fatigue and EPOC. J Expos Sci Environ Epidemiol. 18(3):289-98.

Langstaff JE. (2007). Memorandum to Ozone NAAQS Review Docket (EPA-HQ-OAR-2005-0172). Analysis of Uncertainty in Ozone Population Exposure Modeling. Docket ID No. EPA-HQ-OAR-2005-0172-0174.

Leonard WR, Levy SB, Tarskaia LA, Klimova TM, Fedorova VI, Baltakhinova ME, Krivoshapkin VG, Snodgrass JJ. (2014). Seasonal Variation in Basal Metabolic Rates Among the Yakut (Sakha) of Northeastern Siberia. American Journal of Human Biology. 26(4):437-445.

Ogden CL, Lamb MM, Carroll MD, Flegal KM. (2010). Obesity and Socioeconomic Status in Children and Adolescents: United States, 2005-2008. NCHS Data Brief. Number 51. Available at: https://files.eric.ed.gov/fulltext/ED530165.pdf

Rizzo M. (2008). Memorandum to Sulfur Dioxide Review Docket (EPA-HQ-OAR-2007-0352). Investigation of How Distributions of Hourly Sulfur Dioxide Concentrations Have Changed Over Time in Six Cities. Document ID No. EPA-HQ-OAR-2007-0352-0017

U.S. Census Bureau. (2012). Technical documentation - 2010 Census Summary File 1—Technical Documentation/prepared by the U.S. Census Bureau, Revised 2012. Available at: http://www.census.gov/prod/cen2010/doc/sf1.pdf. Employment Status from the 5-year American Community Survey (ACS) data, 2010 U.S. Census American FactFinder. Available at: http://factfinder2.census.gov/. Commuting times file from U.S. Census data portal (http://dataferrett.census.gov/), Table P31, variables P031001-P031015.

U.S. Census Bureau. (2013). 2013 American Housing Survey (AHS). Available at: https://www.census.gov/programs-surveys/ahs/data/interactive/ahstablecreator.html

U.S. DOT (2012). Bureau of Transportation Statistics, Census Transportation Planning Package, Part 3-The Journey to Work. Available at: http://transtats.bts.gov/

U.S. EPA. (2007). Ozone Population Exposure Analysis for Selected Urban Areas. Office of Air Quality Planning and Standards, Research Triangle Park, NC, EPA-452-R-07-010, July 2007. Available at: http://www.epa.gov/ttn/naaqs/standards/ozone/s_o3_cr_td.html

U.S. EPA. (2008). Risk and Exposure Assessment to Support the Review of the NO_2 Primary National Ambient Air Quality Standard. Office of Air Quality Planning and Standards, Research Triangle Park, NC, EPA-452/R-08-008a, November 2008. Available at: https://www3.epa.gov/ttn/naaqs/standards/nox/s_nox_cr_rea.html

U.S. EPA. (2009). Risk and Exposure Assessment to Support the Review of the SO_2 Primary National Ambient Air Quality Standard. Office of Air Quality Planning and Standards, Research Triangle Park, NC, EPA-452/R-09-007, July 2009. Available at: https://www3.epa.gov/ttn/naaqs/standards/so2/data/200908SO2REAFinalReport.pdf

U.S. EPA. (2010). Quantitative Risk and Exposure Assessment for Carbon Monoxide - Amended. Office of Air Quality Planning and Standards, Research Triangle Park, NC, EPA-452/R-10-006, July 2010. Available at: https://www.epa.gov/naaqs/carbon-monoxide-co-standards-risk-and-exposure-assessments-current-review

U.S. EPA. (2011). Highlights of the Exposure Factors Handbook (Final Report). Office of Research and Development, National Center for Environmental Assessment, Washington, DC, EPA/600/R-10/030, 2011. Available at: https://cfpub.epa.gov/ncea/risk/recordisplay.cfm?deid=221023

U.S. EPA. (2014). Health Risk and Exposure Assessment for Ozone. Office of Air Quality Planning and Standards, Research Triangle Park, NC, EPA-452/R-14-004a, August 2014. Available at: https://www.epa.gov/naaqs/ozone-o3-standards-risk-and-exposure-assessments-current-review

U.S. EPA. (2015). 2011 National Emissions Inventory, version 2 Technical Support Document Documentation. available at: https://www.epa.gov/sites/production/files/2015-10/documents/nei2011v2_tsd_14aug2015.pdf. Data available at: https://www.epa.gov/air-emissions-inventories/2011-national-emissions-inventory-nei-data

U.S. EPA. (2017a). Air Pollutants Exposure Model Documentation (APEX, Version 5) Volume I: User's Guide. Office of Air Quality Planning and Standards, Research Triangle Park, NC, January 2017. EPA-452/R-17-001a. Available at: https://www.epa.gov/fera/apex-user-guides

U.S. EPA. (2017b). Air Pollutants Exposure Model Documentation (APEX, Version 5) Volume II: Technical Support Document. Office of Air Quality Planning and Standards, Research Triangle Park, NC, January 2017. EPA-452/R-17-001b. Available at: https://www.epa.gov/fera/apex-user-guides

U.S. EPA. (2017c). The Consolidated Human Activity Database – Master Version (CHAD-Master). Technical Memorandum. Office of Research and Development, National Exposure Research Laboratory, Research Triangle Park, NC, 27711. In preparation. Previous version (09/15/2014). Available at: https://www.epa.gov/healthresearch/consolidated-human-activity-database-chad-use-human-exposure-and-health-studies-and van Ooijen AMJ, van Marken Lichtenbelt WD, van Steenhovenb AA, Westerterp KR. (2004). Seasonal changes in metabolic and temperature responses to cold air in humans. Physiology & Behavior. 82:545-553.

WHO. (2008). WHO/IPCS Harmonization Project Document No. 6. Part 1: Guidance Document on Characterizing and Communicating Uncertainty in Exposure Assessment. International Programme on Chemical Safety, World Health Organization, Geneva, Switzerland. Available at: http://www.who.int/ipcs/methods/harmonization/areas/exposure/en/

APPENDIX A

SURFACE CHARACTERISTIC VALUES AND METEOROLOGICAL DATA PREPARATION FOR INPUT TO AIR QUALITY MODELING

A.1 Introduction

Air quality dispersion modeling was performed for three study areas to support the SO_2 Risk and Exposure Assessment, including: Fall River, MA; Indianapolis, IN; and Tulsa, OK. Each of the three study areas was modeled for the same three-year period, 2011-2013. National Weather Service (NWS) meteorological data were used as meteorological input to AERMOD (U.S. EPA, 2016a), preprocessed with AERMET (v.16216) (U.S. EPA, 2016b), the meteorological preprocessor for AERMOD.

AERMET requires continuous hourly surface meteorological observations and concurrent twice daily upper air sounding data. The surface and upper air data should be representative of the modeling domain. The NWS and the Federal Aviation Administration (FAA) jointly operate and maintain a network of Automated Surface Observing Systems (ASOS) at airports throughout the U.S. Upper air data are collected by the NWS at 69 stations across the conterminous U.S. Table A-1 and Table A-2 lists the NWS surface and upper air stations selected for each of the study areas. Figure A-1 through Figure A-5 show the locations of the ASOS and upper air stations selected for each study area, relative to emission sources that were modeled.

Table A-1. National Weather Service surface stations.

Study Area	Station	Identifier	WMO (WBAN)	Latitude *(degrees)*	Longitude *(degrees)*	Elevation *(m)*	GMT Offset *(hours)*
Fall River	Providence	PVD	725070 (14765)	41.7225	-71.4325	19	-5
Indianapolis	Indianapolis International Airport	IND	724380 (93819)	39.725170	-86.281680	241	-5
Tulsa	Tulsa R. L. Jones Jr. Airport	RVS	723564 (53908)	36.042441	-95.990166	192	-6

Table A-2. National Weather Service upper air stations.

Study Area	Station	Identifier	WMO (WBAN)	Latitude *(degrees)*	Longitude *(degrees)*	Elevation *(m)*	GMT Offset *(hours)*
Fall River	Chatham, MA	CHH	744940 (14684)	41.67	-69.97	12	-5
Indianapolis	Lincoln, IL	ILX	745600 (04833)	40.15	-89.33	178	-6
Tulsa	Norman, OK	OUN	723560 (13968)	35.23	-97.47	354	-6

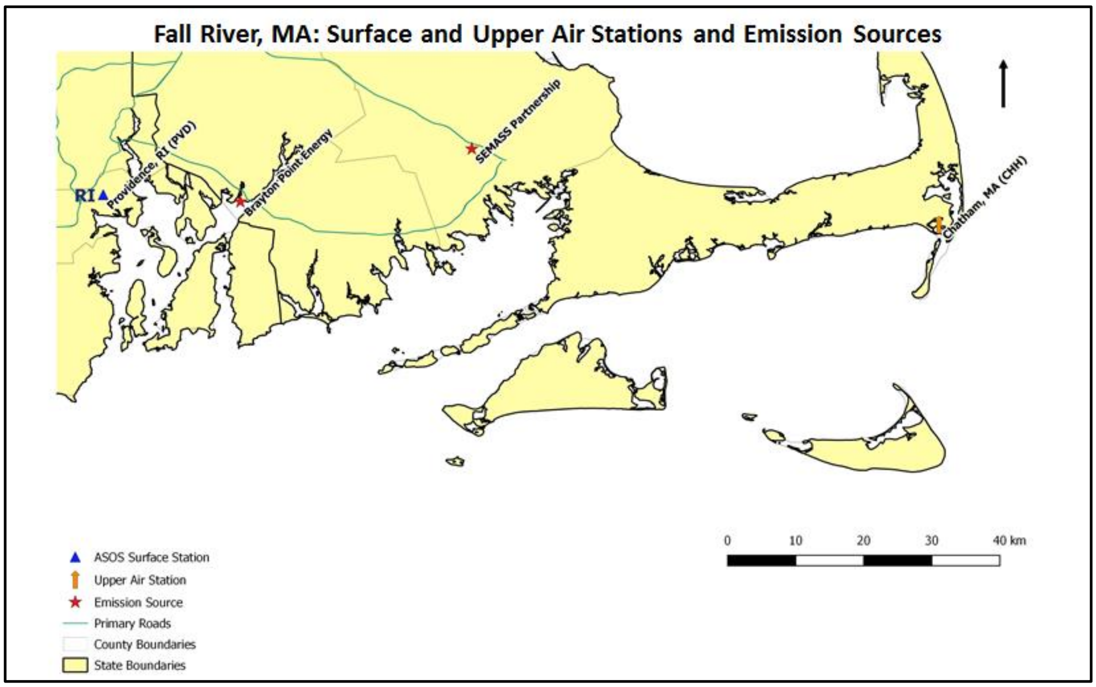


Figure A-1. Location of surface and upper air meteorological stations and emission sources for Fall River, MA.

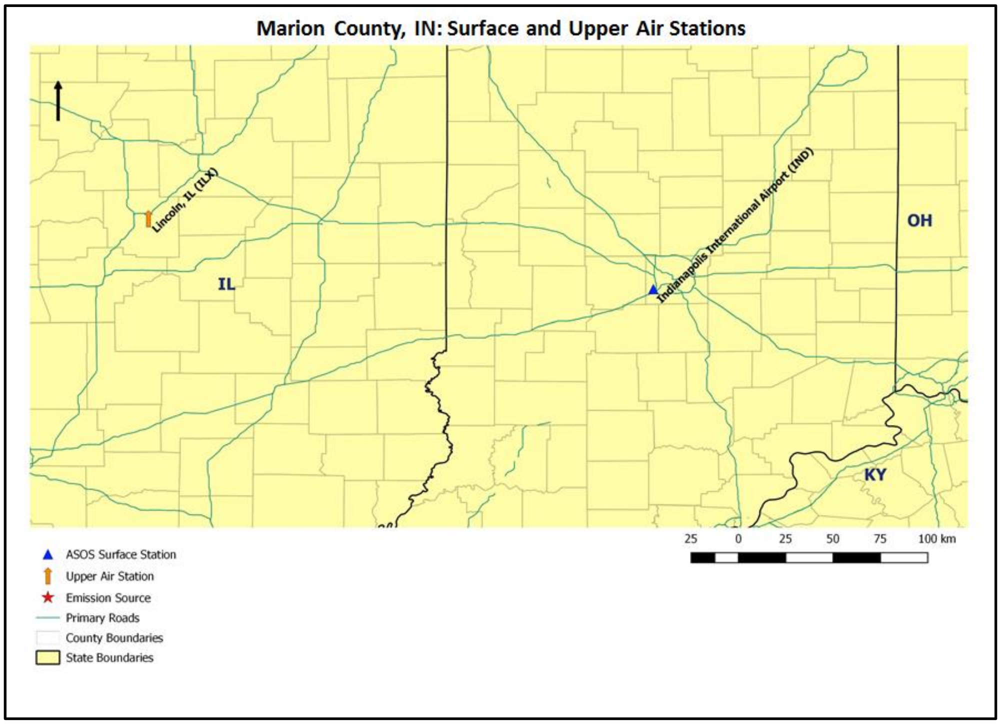


Figure A-2. Location of surface and upper air meteorological stations selected for Indianapolis, IN.

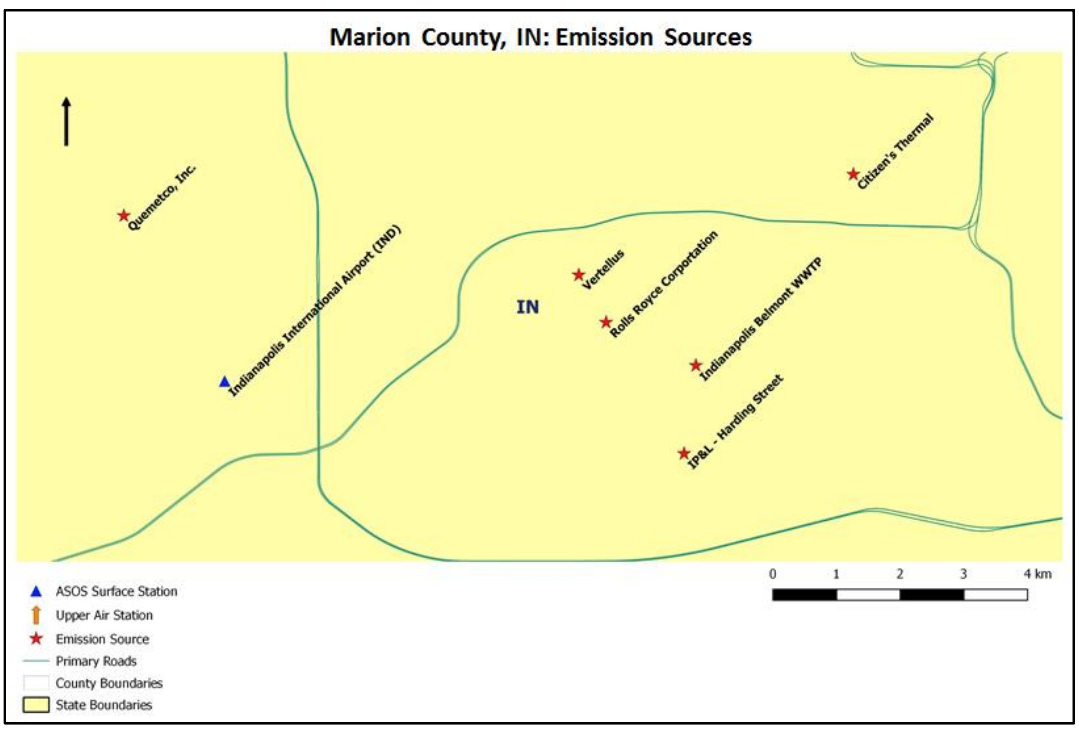


Figure A-3. Location of emission sources for Indianapolis, IN.

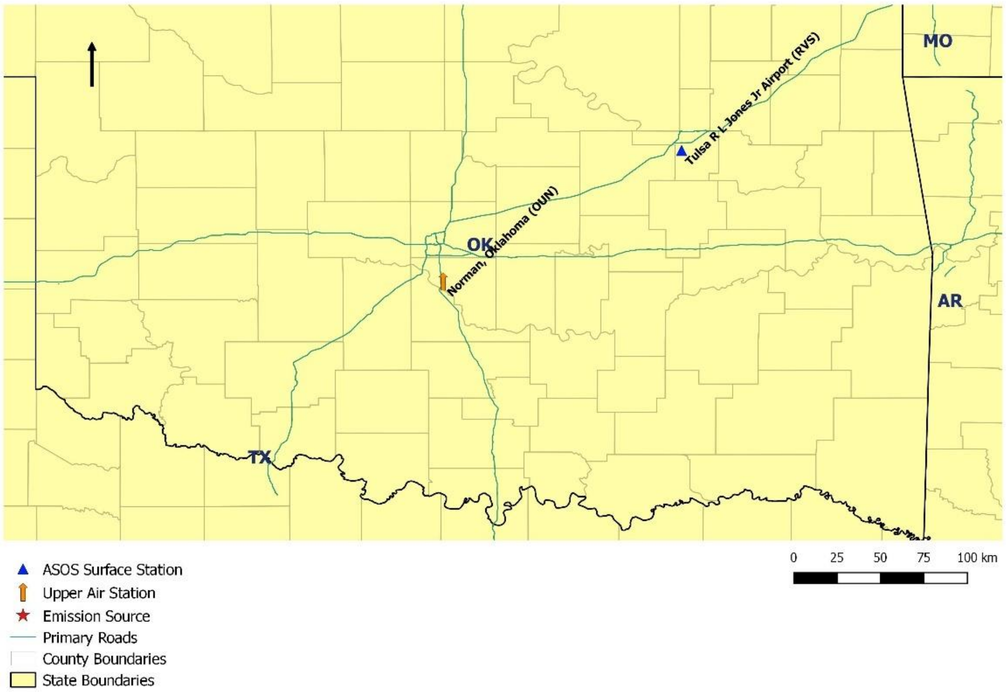


Figure A-4. Location of surface and upper air meteorological stations selected for Tulsa, OK.

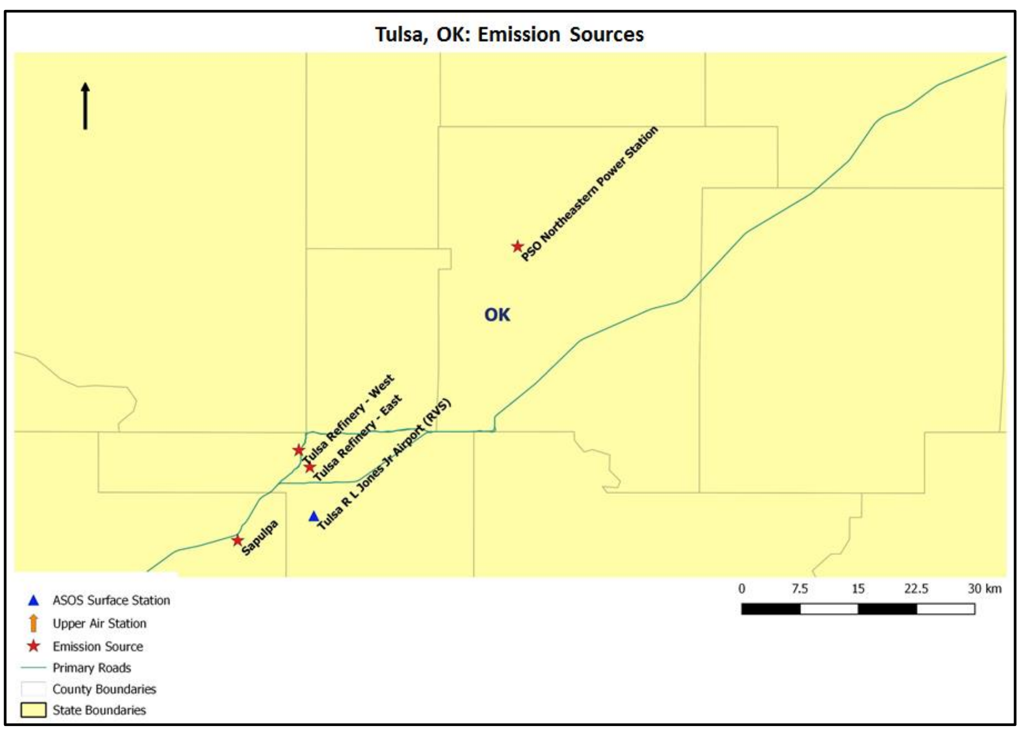


Figure A-5. Location of emission sources for Tulsa, OK.

In addition to surface and upper air meteorological data, AERMET also requires the user to input values of surface albedo, Bowen ratio, and roughness length that are representative of the location where the surface observations are taken. Surface characteristic values were estimated using the AERSURFACE (v.13016) (U.S. EPA, 2013).

The remainder of this document describes the preparation of the meteorological data files input to AERMOD for each of the three study areas. Section A.2 describes the preparation of the surface and upper air data for input to AERMET. Section A.3 describes the estimation of surface characteristic values using AERSURFACE, and Section A.4 describes the AERMET processing with a brief analysis of the AERMET output for each of the study areas.

A.2 Preparation of the Surface and Upper Air Meteorological Data

A.2.1 Surface Data

Three years of surface data for 2011-2013 were downloaded from the Integrated Surface Hourly (ISH) archive maintained by the National Oceanic and Atmospheric Administration (NOAA) National Centers for Environmental Information (NCEI), formerly the National Climatic Data Center (NCDC). The data are accessible for download via File Transfer Protocol (FTP) at ftp://ftp.ncdc.noaa.gov.

A potential concern related to the use of NWS meteorological data for dispersion modeling is the often high incidence of calms and variable wind conditions in the Integrated Surface Hourly (ISH) data. This is due to the implementation of the ASOS program to replace observer-based data beginning in the mid-1990's, and the adoption of the METAR standard for reporting NWS observations in July 1996. Currently, the wind speed and direction used to represent the hour in AERMOD is based on a single two-minute average, usually reported about 10 minutes before the hour. The METAR system reports winds of less than three knots as calm (coded as 0 knots), and winds up to six knots will be reported as variable when the variation in the 2-minute wind direction is more than 60 degrees. This variable wind is reported as a non-zero wind speed with a missing wind direction. The number of calms and variable winds can influence concentration calculations in AERMOD because concentrations are not calculated for calms or variable wind hours. Significant numbers of calm and variable hours may compromise the representativeness of NWS surface data for AERMOD applications. This is especially of concern for applications involving low-level releases since the worst-case dispersion conditions for such sources are associated with low wind speeds, and the hours being discarded as calm or variable are biased toward this condition.

The NCEI maintains a separate archive of 1-minute wind data for each of the ASOS surface stations. These wind data represent 2-minute average wind speeds calculated for each minute of the hour. To reduce the number of calms and missing winds, these wind data were used to calculate hourly average wind speed and direction to replace the standard archive of winds in the ISH dataset. The 1-minute data were processed with AERMINUTE (v.15272) (U.S. EPA, 2015), which calculates the hourly wind speed and wind direction and generates a file formatted for input directly to AERMET, where the ISH wind data are replaced during processing. The NCEI archives the1-minute ASOS wind data as monthly files. Monthly 1-minute data files were downloaded for the 2011-2013 period for each ASOS surface stations listed in Table A-1.

A.2.2 Upper Air Data

Three years (2011-2013) of upper air sounding data were downloaded for each of the upper air stations listed in Table A-2 from the NOAA/ Earth System Research Laboratory (ESRL) Radiosonde Database (https://ruc.noaa.gov/raobs/). The upper air data are archived in the Forecast System Laboratory (FSL) format and maintained by the Global Systems Division, formerly the FSL. Data for each station was downloaded as a separate file as required by AERMET.

A.3 Estimation of Surface Characteristics Using AERSURFACE

As previously stated, surface values for albedo, Bowen ratio, and roughness length were estimated using the AERSURFACE tool. As noted in the AERSURFACE User's Guide (U.S. EPA, 2013), surface characteristics that are input to AERMET should be representative of the location of the meteorological tower. AERSURFACE was run for the location of each of the three ASOS stations using the geographic coordinates of the meteorological towers in Table A-1.

The current version of AERSURFACE utilizes 1992 land cover data from the National Land Cover Database (NLCD) in GeoTIFF format. NLCD data files for the three ASOS stations were downloaded from the Multi-Resolution Land Characteristics consortium website (https://www.mrlc.gov).

AERSURFACE can generate annual, seasonal, or monthly surface characteristic values in a format for input directly into AERMET. Monthly values were generated for each of the locations. To properly interpret some of the land cover categories in the 1992 NLCD data, AERSURFACE requires the user to specify whether or not the location of the weather station is at an airport. All three ASOS stations were specified as airport locations. AERSURFACE also allows for the surface roughness length to be defined by up to 12 wind sectors with a minimum arc of 30 degrees each. For each of the three locations, roughness was estimated for each of 12 sectors, beginning at 0 degrees through 360 degrees (*i.e.*, 0-30, 30-60, 60-90, etc.). The roughness length sectors at each of the three ASOS stations are illustrated in Figure A-6 through

Figure A-8. The sectors extend from the location of the meteorological tower out to 1 km, the distance over which the roughness length is estimated.

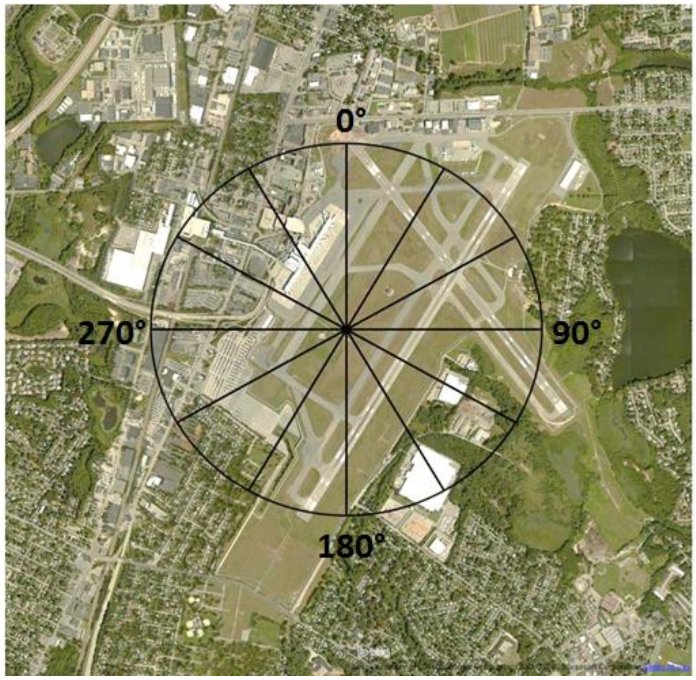


Figure A-6. Surface roughness sectors for Providence Airport (PVD).

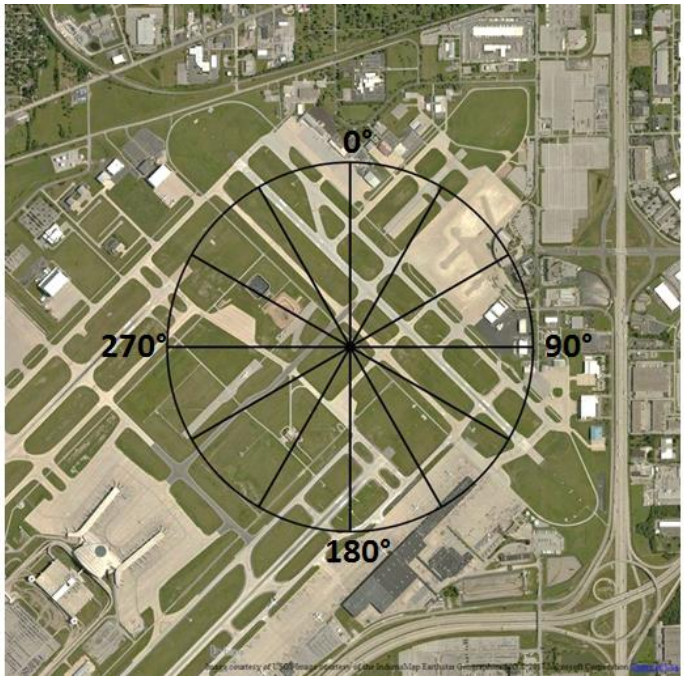


Figure A-7. Surface roughness sectors for Indianapolis International Airport (IND).

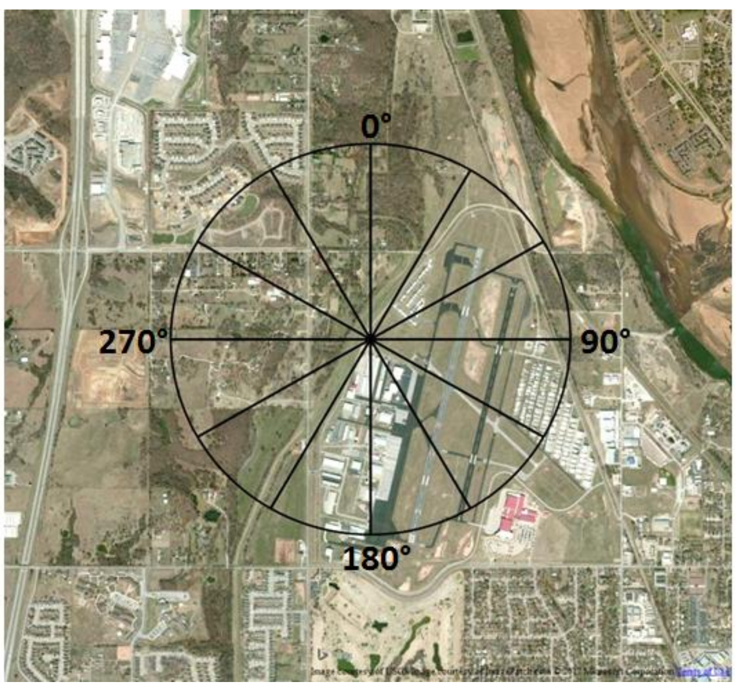


Figure A-8. Surface roughness sectors for Tulsa R. L. Jones Airport (RVS).

Values for the three surface characteristics are defined within AERSURFACE by season but are computed monthly based on the assignment of months to seasons. Monthly values are then rolled up to seasonal or annual values based on the option specified by the user. The user has the option to use default month-to-season assignments or input user-defined assignments. Seasonal surface characteristic values are defined based on five season definitions: spring, summer, autumn, winter with no snow, and winter with continuous snow cover. Note, there are two winter options: 1) winter with no snow (or without continuous snow) on the ground the entire month and 2) winter with continuous snow on ground the entire month.[1] AERSURFACE was run for Tulsa using the default month-to-season assignments, while months were reassigned for both Indianapolis and Fall River. The month-to-season assignments used for each of the three surface stations are shown in Table A-3, along with the seasonal definitions. A month was considered to have continuous snow cover if a snow depth of one inch or more was reported for at least 75% of the days in the month.

Table A-3. AERSURFACE month-to-season assignments.

Station	Winter (continuous snow)	Winter (no snow)	Spring	Summer	Autumn
PVD	Feb (2015 only)	Dec, Jan, Feb, Mar	Apr, May	Jun., Jul, Aug	Sep, Oct, Nov
IND		Dec, Jan, Feb, Mar	Apr, May	Jun., Jul, Aug	Sep, Oct, Nov
RVS		Dec, Jan, Feb	Mar, Apr, May	Jun., Jul, Aug	Sep, Oct, Nov
Seasonal definitions: Winter: Late autumn after frost and harvest, or winter with no snow; Spring: Transitional spring with partial green coverage or short annuals; Summer: Midsummer with lush vegetation; Autumn: Autumn with unharvested cropland					

AERSURFACE also requires information about the climate and surface moisture at the surface station. The climate at the station location is categorized as either arid or non-arid. Each of the three surface station locations was categorized as non-arid in AERSURFACE. Surface moisture is based on precipitation amounts and is categorized as either wet, average, or dry. For the three surface stations, 2010 local climatological data from the NCEI was used to look at 30 years (1981-2010) of monthly precipitation. The 30th and 70th percentiles of precipitation amounts were calculated for each of 12 months (Jan. – Dec.) based on the 30-year period. The precipitation amount for each month in 2011-2013 was then compared to the 30th and 70th percentiles for the corresponding month. Months during which precipitation was greater than the 70th percentile were considered wet while months that were less than the 30th percentile were considered dry. Months within the 30th and 70th percentile range were considered average. AERSURFACE was run for each moisture condition to obtain monthly values for wet, dry, and average conditions. Using the AERSURFACE output for each of the three moisture categories, a

[1] For many of the land cover categories in the 1992 NLCD classification scheme, the designation of winter with continuous snow on the ground would tend to increase wintertime albedo (reflectivity) and decrease wintertime Bowen ratio (sensible to latent heat flux) and surface roughness compared to the winter with no snow or without continuous snow designation.

separate set of monthly surface characteristics was compiled for each of the three years for input to AERMET. The monthly categorization of the surface moisture at each of the locations is shown in Table A-4. The resulting surface characteristic values input to AERMET, by sector, month, and year, are listed in Table A-6 through Table A-8 at the end of this document.

Table A-4. Monthly surface moisture categorizations.

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
PVD												
2011	Avg	Wet	Dry	Wet	Avg	Wet	Wet	Wet	Wet	Wet	Wet	Avg
2012	Avg	Dry	Dry	Avg	Wet	Wet	Avg	Wet	Wet	Wet	Dry	Wet
2013	Dry	Wet	Dry	Dry	Avg	Wet	Avg	Wet	Wet	Dry	Wet	Wet
IND												
2011	Wet	Wet	Wet	Wet	Wet	Wet	Dry	Dry	Wet	Wet	Wet	Wet
2012	Wet	Avg	Wet	Avg	Dry	Dry	Dry	Wet	Wet	Wet	Dry	Avg
2013	Wet	Wet	Wet	Wet	Wet	Wet	Dry	Dry	Wet	Wet	Wet	Wet
RVS *(Moisture conditions at RVS are based on precipitation data from Tulsa International Airport, TUL)*												
2011	Dry	Wet	Dry	Wet	Dry	Dry	Dry	Wet	Dry	Dry	Wet	Avg
2012	Dry	Avg	Wet	Avg	Dry	Wet	Dry	Wet	Dry	Avg	Dry	Dry
2013	Wet	Wet	Dry	Avg	Avg	Dry	Wet	Wet	Dry	Wet	Avg	Avg

A.4 AERMET Processing

The meteorological data files (upper air, ISH data, and 1-minute hourly averaged wind data) for each station were processed in AERMET. Each year was processed separately using the monthly surface characteristics specific to each year. AERMET processes the meteorological data in three “Stages.” Stage 1 reads in the upper air and ISH data files and performs an initial QA on the values. Stage 2 reads the 1-minute averaged wind data and merges the three data sets into a single file. Stage 3 performs data replacements and substitutions as specified by the user, computes the boundary layer parameters, and generates data files formatted for input to AERMOD. Surface characteristics were input during Stage 3. When 1-minute hourly averaged winds were available, those winds were used for the hour, while all other surface data are from the ISH data (temperature, cloud cover, precipitation, etc.).

Table A-5 shows the percentage of calm and missing winds in the AERMET output for the combined three years (2011-2013) for each of the surface stations. These values take into account the replacement of the ISH wind data with the 1-minute hourly averaged wind data during the AERMET Stage 3 processing. Figure A-9 through Figure A-11 are wind roses generated from the 2011-2013 surface data files output by AERMET for three surface stations.

Table A-5. Percent calm and missing winds in AERMET surface file.

Station	% Calm	% Missing
PVD	0.49	0.06
IND	0.37	0.10
RVS	3.90	0.22

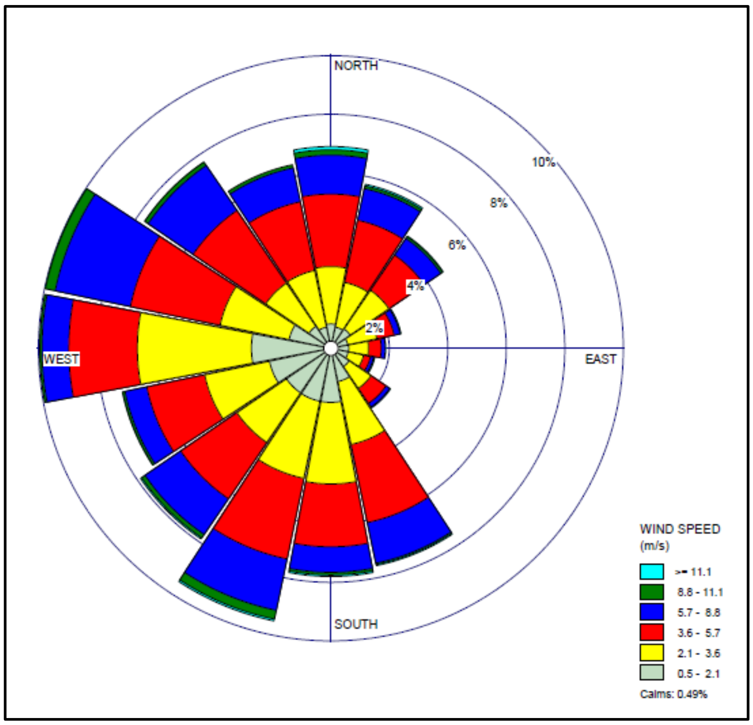


Figure A-9. Wind rose for Providence Airport (PVD), 2011-2013 (direction blowing from).

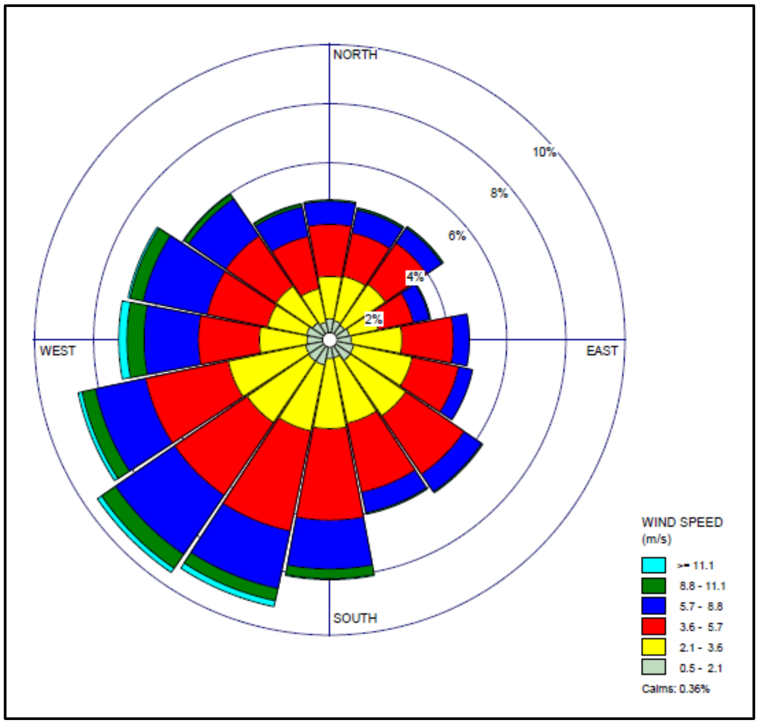


Figure A-10. Wind rose for Indianapolis International Airport (IND), 2011-2013 (direction blowing from).

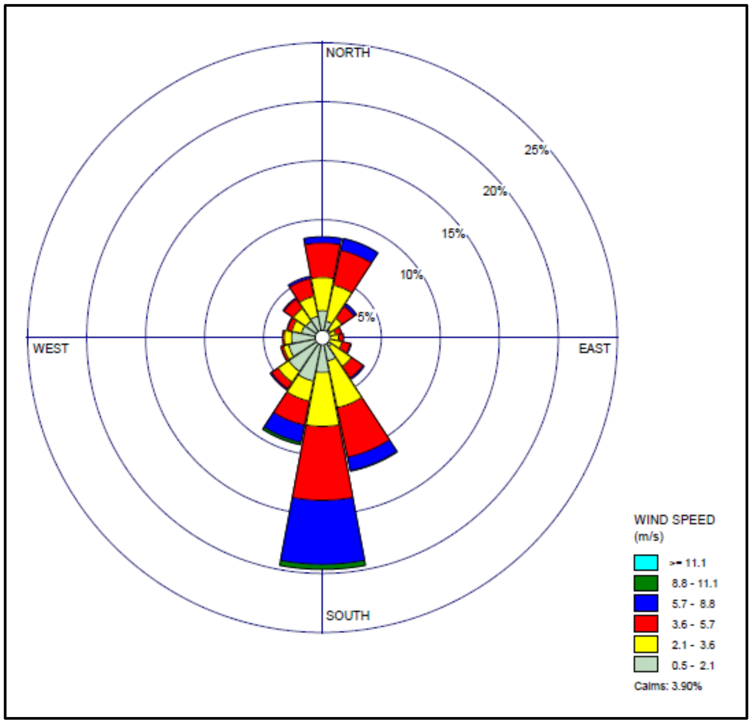


Figure A-11. Wind rose for Tulsa R. L. Jones Jr. Airport (RVS), 2011-2013 (direction blowing from).

Table A-6. Surface characteristics for Providence Airport (PVD) by month and year.

Station = PVD		2011			2012			2013		
Month	Sector (degrees)	Albedo	Bowen Ratio	Roughness (m)	Albedo	Bowen Ratio	Roughness (m)	Albedo	Bowen Ratio	Roughness (m)
Jan	0-30	0.16	0.64	0.023	0.16	0.64	0.023	0.16	1.24	0.023
Jan	30-60	0.16	0.64	0.022	0.16	0.64	0.022	0.16	1.24	0.022
Jan	60-90	0.16	0.64	0.026	0.16	0.64	0.026	0.16	1.24	0.026
Jan	90-120	0.16	0.64	0.036	0.16	0.64	0.036	0.16	1.24	0.036
Jan	120-150	0.16	0.64	0.041	0.16	0.64	0.041	0.16	1.24	0.041
Jan	150-180	0.16	0.64	0.027	0.16	0.64	0.027	0.16	1.24	0.027
Jan	180-210	0.16	0.64	0.018	0.16	0.64	0.018	0.16	1.24	0.018
Jan	210-240	0.16	0.64	0.038	0.16	0.64	0.038	0.16	1.24	0.038
Jan	240-270	0.16	0.64	0.038	0.16	0.64	0.038	0.16	1.24	0.038
Jan	270-300	0.16	0.64	0.053	0.16	0.64	0.053	0.16	1.24	0.053
Jan	300-330	0.16	0.64	0.081	0.16	0.64	0.081	0.16	1.24	0.081
Jan	330-360	0.16	0.64	0.030	0.16	0.64	0.030	0.16	1.24	0.030
Feb	0-30	0.16	0.40	0.023	0.16	1.24	0.023	0.16	0.40	0.023
Feb	30-60	0.16	0.40	0.022	0.16	1.24	0.022	0.16	0.40	0.022
Feb	60-90	0.16	0.40	0.026	0.16	1.24	0.026	0.16	0.40	0.026
Feb	90-120	0.16	0.40	0.036	0.16	1.24	0.036	0.16	0.40	0.036
Feb	120-150	0.16	0.40	0.041	0.16	1.24	0.041	0.16	0.40	0.041
Feb	150-180	0.16	0.40	0.027	0.16	1.24	0.027	0.16	0.40	0.027
Feb	180-210	0.16	0.40	0.018	0.16	1.24	0.018	0.16	0.40	0.018
Feb	210-240	0.16	0.40	0.038	0.16	1.24	0.038	0.16	0.40	0.038
Feb	240-270	0.16	0.40	0.038	0.16	1.24	0.038	0.16	0.40	0.038
Feb	270-300	0.16	0.40	0.053	0.16	1.24	0.053	0.16	0.40	0.053
Feb	300-330	0.16	0.40	0.081	0.16	1.24	0.081	0.16	0.40	0.081
Feb	330-360	0.16	0.40	0.030	0.16	1.24	0.030	0.16	0.40	0.030
Mar	0-30	0.16	1.24	0.023	0.16	1.24	0.023	0.16	1.24	0.023
Mar	30-60	0.16	1.24	0.022	0.16	1.24	0.022	0.16	1.24	0.022
Mar	60-90	0.16	1.24	0.026	0.16	1.24	0.026	0.16	1.24	0.026
Mar	90-120	0.16	1.24	0.036	0.16	1.24	0.036	0.16	1.24	0.036
Mar	120-150	0.16	1.24	0.041	0.16	1.24	0.041	0.16	1.24	0.041
Mar	150-180	0.16	1.24	0.027	0.16	1.24	0.027	0.16	1.24	0.027
Mar	180-210	0.16	1.24	0.018	0.16	1.24	0.018	0.16	1.24	0.018
Mar	210-240	0.16	1.24	0.038	0.16	1.24	0.038	0.16	1.24	0.038
Mar	240-270	0.16	1.24	0.038	0.16	1.24	0.038	0.16	1.24	0.038
Mar	270-300	0.16	1.24	0.053	0.16	1.24	0.053	0.16	1.24	0.053
Mar	300-330	0.16	1.24	0.081	0.16	1.24	0.081	0.16	1.24	0.081
Mar	330-360	0.16	1.24	0.030	0.16	1.24	0.030	0.16	1.24	0.030
Apr	0-30	0.15	0.37	0.029	0.15	0.53	0.029	0.15	1.05	0.029
Apr	30-60	0.15	0.37	0.029	0.15	0.53	0.029	0.15	1.05	0.029

Station = PVD		2011			2012			2013		
Month	Sector (degrees)	Albedo	Bowen Ratio	Roughness (m)	Albedo	Bowen Ratio	Roughness (m)	Albedo	Bowen Ratio	Roughness (m)
Apr	60-90	0.15	0.37	0.034	0.15	0.53	0.034	0.15	1.05	0.034
Apr	90-120	0.15	0.37	0.047	0.15	0.53	0.047	0.15	1.05	0.047
Apr	120-150	0.15	0.37	0.052	0.15	0.53	0.052	0.15	1.05	0.052
Apr	150-180	0.15	0.37	0.036	0.15	0.53	0.036	0.15	1.05	0.036
Apr	180-210	0.15	0.37	0.025	0.15	0.53	0.025	0.15	1.05	0.025
Apr	210-240	0.15	0.37	0.051	0.15	0.53	0.051	0.15	1.05	0.051
Apr	240-270	0.15	0.37	0.045	0.15	0.53	0.045	0.15	1.05	0.045
Apr	270-300	0.15	0.37	0.062	0.15	0.53	0.062	0.15	1.05	0.062
Apr	300-330	0.15	0.37	0.088	0.15	0.53	0.088	0.15	1.05	0.088
Apr	330-360	0.15	0.37	0.037	0.15	0.53	0.037	0.15	1.05	0.037
May	0-30	0.15	0.53	0.029	0.15	0.37	0.029	0.15	0.53	0.029
May	30-60	0.15	0.53	0.029	0.15	0.37	0.029	0.15	0.53	0.029
May	60-90	0.15	0.53	0.034	0.15	0.37	0.034	0.15	0.53	0.034
May	90-120	0.15	0.53	0.047	0.15	0.37	0.047	0.15	0.53	0.047
May	120-150	0.15	0.53	0.052	0.15	0.37	0.052	0.15	0.53	0.052
May	150-180	0.15	0.53	0.036	0.15	0.37	0.036	0.15	0.53	0.036
May	180-210	0.15	0.53	0.025	0.15	0.37	0.025	0.15	0.53	0.025
May	210-240	0.15	0.53	0.051	0.15	0.37	0.051	0.15	0.53	0.051
May	240-270	0.15	0.53	0.045	0.15	0.37	0.045	0.15	0.53	0.045
May	270-300	0.15	0.53	0.062	0.15	0.37	0.062	0.15	0.53	0.062
May	300-330	0.15	0.53	0.088	0.15	0.37	0.088	0.15	0.53	0.088
May	330-360	0.15	0.53	0.037	0.15	0.37	0.037	0.15	0.53	0.037
Jun	0-30	0.15	0.36	0.035	0.15	0.36	0.035	0.15	0.36	0.035
Jun	30-60	0.15	0.36	0.035	0.15	0.36	0.035	0.15	0.36	0.035
Jun	60-90	0.15	0.36	0.040	0.15	0.36	0.040	0.15	0.36	0.040
Jun	90-120	0.15	0.36	0.056	0.15	0.36	0.056	0.15	0.36	0.056
Jun	120-150	0.15	0.36	0.061	0.15	0.36	0.061	0.15	0.36	0.061
Jun	150-180	0.15	0.36	0.043	0.15	0.36	0.043	0.15	0.36	0.043
Jun	180-210	0.15	0.36	0.031	0.15	0.36	0.031	0.15	0.36	0.031
Jun	210-240	0.15	0.36	0.059	0.15	0.36	0.059	0.15	0.36	0.059
Jun	240-270	0.15	0.36	0.050	0.15	0.36	0.050	0.15	0.36	0.050
Jun	270-300	0.15	0.36	0.068	0.15	0.36	0.068	0.15	0.36	0.068
Jun	300-330	0.15	0.36	0.094	0.15	0.36	0.094	0.15	0.36	0.094
Jun	330-360	0.15	0.36	0.042	0.15	0.36	0.042	0.15	0.36	0.042
Jul	0-30	0.15	0.36	0.035	0.15	0.49	0.035	0.15	0.49	0.035
Jul	30-60	0.15	0.36	0.035	0.15	0.49	0.035	0.15	0.49	0.035
Jul	60-90	0.15	0.36	0.040	0.15	0.49	0.040	0.15	0.49	0.040
Jul	90-120	0.15	0.36	0.056	0.15	0.49	0.056	0.15	0.49	0.056
Jul	120-150	0.15	0.36	0.061	0.15	0.49	0.061	0.15	0.49	0.061

Station = PVD		2011			2012			2013		
Month	Sector (degrees)	Albedo	Bowen Ratio	Roughness (m)	Albedo	Bowen Ratio	Roughness (m)	Albedo	Bowen Ratio	Roughness (m)
Jul	150-180	0.15	0.36	0.043	0.15	0.49	0.043	0.15	0.49	0.043
Jul	180-210	0.15	0.36	0.031	0.15	0.49	0.031	0.15	0.49	0.031
Jul	210-240	0.15	0.36	0.059	0.15	0.49	0.059	0.15	0.49	0.059
Jul	240-270	0.15	0.36	0.050	0.15	0.49	0.050	0.15	0.49	0.050
Jul	270-300	0.15	0.36	0.068	0.15	0.49	0.068	0.15	0.49	0.068
Jul	300-330	0.15	0.36	0.094	0.15	0.49	0.094	0.15	0.49	0.094
Jul	330-360	0.15	0.36	0.042	0.15	0.49	0.042	0.15	0.49	0.042
Aug	0-30	0.15	0.36	0.035	0.15	0.36	0.035	0.15	0.36	0.035
Aug	30-60	0.15	0.36	0.035	0.15	0.36	0.035	0.15	0.36	0.035
Aug	60-90	0.15	0.36	0.040	0.15	0.36	0.040	0.15	0.36	0.040
Aug	90-120	0.15	0.36	0.056	0.15	0.36	0.056	0.15	0.36	0.056
Aug	120-150	0.15	0.36	0.061	0.15	0.36	0.061	0.15	0.36	0.061
Aug	150-180	0.15	0.36	0.043	0.15	0.36	0.043	0.15	0.36	0.043
Aug	180-210	0.15	0.36	0.031	0.15	0.36	0.031	0.15	0.36	0.031
Aug	210-240	0.15	0.36	0.059	0.15	0.36	0.059	0.15	0.36	0.059
Aug	240-270	0.15	0.36	0.050	0.15	0.36	0.050	0.15	0.36	0.050
Aug	270-300	0.15	0.36	0.068	0.15	0.36	0.068	0.15	0.36	0.068
Aug	300-330	0.15	0.36	0.094	0.15	0.36	0.094	0.15	0.36	0.094
Aug	330-360	0.15	0.36	0.042	0.15	0.36	0.042	0.15	0.36	0.042
Sep	0-30	0.15	0.40	0.029	0.15	0.40	0.029	0.15	0.40	0.029
Sep	30-60	0.15	0.40	0.029	0.15	0.40	0.029	0.15	0.40	0.029
Sep	60-90	0.15	0.40	0.034	0.15	0.40	0.034	0.15	0.40	0.034
Sep	90-120	0.15	0.40	0.048	0.15	0.40	0.048	0.15	0.40	0.048
Sep	120-150	0.15	0.40	0.053	0.15	0.40	0.053	0.15	0.40	0.053
Sep	150-180	0.15	0.40	0.036	0.15	0.40	0.036	0.15	0.40	0.036
Sep	180-210	0.15	0.40	0.025	0.15	0.40	0.025	0.15	0.40	0.025
Sep	210-240	0.15	0.40	0.051	0.15	0.40	0.051	0.15	0.40	0.051
Sep	240-270	0.15	0.40	0.045	0.15	0.40	0.045	0.15	0.40	0.045
Sep	270-300	0.15	0.40	0.062	0.15	0.40	0.062	0.15	0.40	0.062
Sep	300-330	0.15	0.40	0.088	0.15	0.40	0.088	0.15	0.40	0.088
Sep	330-360	0.15	0.40	0.037	0.15	0.40	0.037	0.15	0.40	0.037
Oct	0-30	0.15	0.40	0.029	0.15	0.40	0.029	0.15	1.24	0.029
Oct	30-60	0.15	0.40	0.029	0.15	0.40	0.029	0.15	1.24	0.029
Oct	60-90	0.15	0.40	0.034	0.15	0.40	0.034	0.15	1.24	0.034
Oct	90-120	0.15	0.40	0.048	0.15	0.40	0.048	0.15	1.24	0.048
Oct	120-150	0.15	0.40	0.053	0.15	0.40	0.053	0.15	1.24	0.053
Oct	150-180	0.15	0.40	0.036	0.15	0.40	0.036	0.15	1.24	0.036
Oct	180-210	0.15	0.40	0.025	0.15	0.40	0.025	0.15	1.24	0.025
Oct	210-240	0.15	0.40	0.051	0.15	0.40	0.051	0.15	1.24	0.051

Station = PVD		2011			2012			2013		
Month	Sector (degrees)	Albedo	Bowen Ratio	Roughness (m)	Albedo	Bowen Ratio	Roughness (m)	Albedo	Bowen Ratio	Roughness (m)
Oct	240-270	0.15	0.40	0.045	0.15	0.40	0.045	0.15	1.24	0.045
Oct	270-300	0.15	0.40	0.062	0.15	0.40	0.062	0.15	1.24	0.062
Oct	300-330	0.15	0.40	0.088	0.15	0.40	0.088	0.15	1.24	0.088
Oct	330-360	0.15	0.40	0.037	0.15	0.40	0.037	0.15	1.24	0.037
Nov	0-30	0.15	0.40	0.029	0.15	1.24	0.029	0.15	0.40	0.029
Nov	30-60	0.15	0.40	0.029	0.15	1.24	0.029	0.15	0.40	0.029
Nov	60-90	0.15	0.40	0.034	0.15	1.24	0.034	0.15	0.40	0.034
Nov	90-120	0.15	0.40	0.048	0.15	1.24	0.048	0.15	0.40	0.048
Nov	120-150	0.15	0.40	0.053	0.15	1.24	0.053	0.15	0.40	0.053
Nov	150-180	0.15	0.40	0.036	0.15	1.24	0.036	0.15	0.40	0.036
Nov	180-210	0.15	0.40	0.025	0.15	1.24	0.025	0.15	0.40	0.025
Nov	210-240	0.15	0.40	0.051	0.15	1.24	0.051	0.15	0.40	0.051
Nov	240-270	0.15	0.40	0.045	0.15	1.24	0.045	0.15	0.40	0.045
Nov	270-300	0.15	0.40	0.062	0.15	1.24	0.062	0.15	0.40	0.062
Nov	300-330	0.15	0.40	0.088	0.15	1.24	0.088	0.15	0.40	0.088
Nov	330-360	0.15	0.40	0.037	0.15	1.24	0.037	0.15	0.40	0.037
Dec	0-30	0.16	0.64	0.023	0.16	0.40	0.023	0.16	0.40	0.023
Dec	30-60	0.16	0.64	0.022	0.16	0.40	0.022	0.16	0.40	0.022
Dec	60-90	0.16	0.64	0.026	0.16	0.40	0.026	0.16	0.40	0.026
Dec	90-120	0.16	0.64	0.036	0.16	0.40	0.036	0.16	0.40	0.036
Dec	120-150	0.16	0.64	0.041	0.16	0.40	0.041	0.16	0.40	0.041
Dec	150-180	0.16	0.64	0.027	0.16	0.40	0.027	0.16	0.40	0.027
Dec	180-210	0.16	0.64	0.018	0.16	0.40	0.018	0.16	0.40	0.018
Dec	210-240	0.16	0.64	0.038	0.16	0.40	0.038	0.16	0.40	0.038
Dec	240-270	0.16	0.64	0.038	0.16	0.40	0.038	0.16	0.40	0.038
Dec	270-300	0.16	0.64	0.053	0.16	0.40	0.053	0.16	0.40	0.053
Dec	300-330	0.16	0.64	0.081	0.16	0.40	0.081	0.16	0.40	0.081
Dec	330-360	0.16	0.64	0.030	0.16	0.40	0.030	0.16	0.40	0.030

Table A-7. Surface characteristics for Indianapolis Int'l (IND) by month and year.

Station = IND		2011			2012			2013		
Month	Sector (degrees)	Albedo	Bowen Ratio	Roughness (m)	Albedo	Bowen Ratio	Roughness (m)	Albedo	Bowen Ratio	Roughness (m)
Jan	0-30	0.18	0.52	0.032	0.18	0.52	0.032	0.18	0.52	0.032
Jan	30-60	0.18	0.52	0.033	0.18	0.52	0.033	0.18	0.52	0.033
Jan	60-90	0.18	0.52	0.046	0.18	0.52	0.046	0.18	0.52	0.046
Jan	90-120	0.18	0.52	0.030	0.18	0.52	0.030	0.18	0.52	0.030
Jan	120-150	0.18	0.52	0.031	0.18	0.52	0.031	0.18	0.52	0.031
Jan	150-180	0.18	0.52	0.040	0.18	0.52	0.040	0.18	0.52	0.040
Jan	180-210	0.18	0.52	0.027	0.18	0.52	0.027	0.18	0.52	0.027
Jan	210-240	0.18	0.52	0.016	0.18	0.52	0.016	0.18	0.52	0.016
Jan	240-270	0.18	0.52	0.022	0.18	0.52	0.022	0.18	0.52	0.022
Jan	270-300	0.18	0.52	0.022	0.18	0.52	0.022	0.18	0.52	0.022
Jan	300-330	0.18	0.52	0.019	0.18	0.52	0.019	0.18	0.52	0.019
Jan	330-360	0.18	0.52	0.041	0.18	0.52	0.041	0.18	0.52	0.041
Feb	0-30	0.18	0.52	0.032	0.18	0.89	0.032	0.18	0.52	0.032
Feb	30-60	0.18	0.52	0.033	0.18	0.89	0.033	0.18	0.52	0.033
Feb	60-90	0.18	0.52	0.046	0.18	0.89	0.046	0.18	0.52	0.046
Feb	90-120	0.18	0.52	0.030	0.18	0.89	0.030	0.18	0.52	0.030
Feb	120-150	0.18	0.52	0.031	0.18	0.89	0.031	0.18	0.52	0.031
Feb	150-180	0.18	0.52	0.040	0.18	0.89	0.040	0.18	0.52	0.040
Feb	180-210	0.18	0.52	0.027	0.18	0.89	0.027	0.18	0.52	0.027
Feb	210-240	0.18	0.52	0.016	0.18	0.89	0.016	0.18	0.52	0.016
Feb	240-270	0.18	0.52	0.022	0.18	0.89	0.022	0.18	0.52	0.022
Feb	270-300	0.18	0.52	0.022	0.18	0.89	0.022	0.18	0.52	0.022
Feb	300-330	0.18	0.52	0.019	0.18	0.89	0.019	0.18	0.52	0.019
Feb	330-360	0.18	0.52	0.041	0.18	0.89	0.041	0.18	0.52	0.041
Mar	0-30	0.15	0.36	0.038	0.15	0.36	0.038	0.15	0.36	0.038
Mar	30-60	0.15	0.36	0.039	0.15	0.36	0.039	0.15	0.36	0.039
Mar	60-90	0.15	0.36	0.051	0.15	0.36	0.051	0.15	0.36	0.051
Mar	90-120	0.15	0.36	0.036	0.15	0.36	0.036	0.15	0.36	0.036
Mar	120-150	0.15	0.36	0.038	0.15	0.36	0.038	0.15	0.36	0.038
Mar	150-180	0.15	0.36	0.046	0.15	0.36	0.046	0.15	0.36	0.046
Mar	180-210	0.15	0.36	0.034	0.15	0.36	0.034	0.15	0.36	0.034
Mar	210-240	0.15	0.36	0.022	0.15	0.36	0.022	0.15	0.36	0.022
Mar	240-270	0.15	0.36	0.029	0.15	0.36	0.029	0.15	0.36	0.029
Mar	270-300	0.15	0.36	0.028	0.15	0.36	0.028	0.15	0.36	0.028
Mar	300-330	0.15	0.36	0.025	0.15	0.36	0.025	0.15	0.36	0.025
Mar	330-360	0.15	0.36	0.046	0.15	0.36	0.046	0.15	0.36	0.046
Apr	0-30	0.15	0.36	0.038	0.15	0.53	0.038	0.15	0.36	0.038
Apr	30-60	0.15	0.36	0.039	0.15	0.53	0.039	0.15	0.36	0.039

Station = IND		2011			2012			2013		
Month	Sector (degrees)	Albedo	Bowen Ratio	Roughness (m)	Albedo	Bowen Ratio	Roughness (m)	Albedo	Bowen Ratio	Roughness (m)
Apr	60-90	0.15	0.36	0.051	0.15	0.53	0.051	0.15	0.36	0.051
Apr	90-120	0.15	0.36	0.036	0.15	0.53	0.036	0.15	0.36	0.036
Apr	120-150	0.15	0.36	0.038	0.15	0.53	0.038	0.15	0.36	0.038
Apr	150-180	0.15	0.36	0.046	0.15	0.53	0.046	0.15	0.36	0.046
Apr	180-210	0.15	0.36	0.034	0.15	0.53	0.034	0.15	0.36	0.034
Apr	210-240	0.15	0.36	0.022	0.15	0.53	0.022	0.15	0.36	0.022
Apr	240-270	0.15	0.36	0.029	0.15	0.53	0.029	0.15	0.36	0.029
Apr	270-300	0.15	0.36	0.028	0.15	0.53	0.028	0.15	0.36	0.028
Apr	300-330	0.15	0.36	0.025	0.15	0.53	0.025	0.15	0.36	0.025
Apr	330-360	0.15	0.36	0.046	0.15	0.53	0.046	0.15	0.36	0.046
May	0-30	0.15	0.36	0.038	0.15	1.47	0.038	0.15	0.36	0.038
May	30-60	0.15	0.36	0.039	0.15	1.47	0.039	0.15	0.36	0.039
May	60-90	0.15	0.36	0.051	0.15	1.47	0.051	0.15	0.36	0.051
May	90-120	0.15	0.36	0.036	0.15	1.47	0.036	0.15	0.36	0.036
May	120-150	0.15	0.36	0.038	0.15	1.47	0.038	0.15	0.36	0.038
May	150-180	0.15	0.36	0.046	0.15	1.47	0.046	0.15	0.36	0.046
May	180-210	0.15	0.36	0.034	0.15	1.47	0.034	0.15	0.36	0.034
May	210-240	0.15	0.36	0.022	0.15	1.47	0.022	0.15	0.36	0.022
May	240-270	0.15	0.36	0.029	0.15	1.47	0.029	0.15	0.36	0.029
May	270-300	0.15	0.36	0.028	0.15	1.47	0.028	0.15	0.36	0.028
May	300-330	0.15	0.36	0.025	0.15	1.47	0.025	0.15	0.36	0.025
May	330-360	0.15	0.36	0.046	0.15	1.47	0.046	0.15	0.36	0.046
Jun	0-30	0.18	0.44	0.045	0.18	1.76	0.045	0.18	0.44	0.045
Jun	30-60	0.18	0.44	0.045	0.18	1.76	0.045	0.18	0.44	0.045
Jun	60-90	0.18	0.44	0.056	0.18	1.76	0.056	0.18	0.44	0.056
Jun	90-120	0.18	0.44	0.046	0.18	1.76	0.046	0.18	0.44	0.046
Jun	120-150	0.18	0.44	0.048	0.18	1.76	0.048	0.18	0.44	0.048
Jun	150-180	0.18	0.44	0.063	0.18	1.76	0.063	0.18	0.44	0.063
Jun	180-210	0.18	0.44	0.051	0.18	1.76	0.051	0.18	0.44	0.051
Jun	210-240	0.18	0.44	0.040	0.18	1.76	0.040	0.18	0.44	0.040
Jun	240-270	0.18	0.44	0.043	0.18	1.76	0.043	0.18	0.44	0.043
Jun	270-300	0.18	0.44	0.042	0.18	1.76	0.042	0.18	0.44	0.042
Jun	300-330	0.18	0.44	0.032	0.18	1.76	0.032	0.18	0.44	0.032
Jun	330-360	0.18	0.44	0.055	0.18	1.76	0.055	0.18	0.44	0.055
Jul	0-30	0.18	1.76	0.045	0.18	1.76	0.045	0.18	1.76	0.045
Jul	30-60	0.18	1.76	0.045	0.18	1.76	0.045	0.18	1.76	0.045
Jul	60-90	0.18	1.76	0.056	0.18	1.76	0.056	0.18	1.76	0.056
Jul	90-120	0.18	1.76	0.046	0.18	1.76	0.046	0.18	1.76	0.046
Jul	120-150	0.18	1.76	0.048	0.18	1.76	0.048	0.18	1.76	0.048

Station = IND		2011			2012			2013		
Month	Sector (degrees)	Albedo	Bowen Ratio	Roughness (m)	Albedo	Bowen Ratio	Roughness (m)	Albedo	Bowen Ratio	Roughness (m)
Jul	150-180	0.18	1.76	0.063	0.18	1.76	0.063	0.18	1.76	0.063
Jul	180-210	0.18	1.76	0.051	0.18	1.76	0.051	0.18	1.76	0.051
Jul	210-240	0.18	1.76	0.040	0.18	1.76	0.040	0.18	1.76	0.040
Jul	240-270	0.18	1.76	0.043	0.18	1.76	0.043	0.18	1.76	0.043
Jul	270-300	0.18	1.76	0.042	0.18	1.76	0.042	0.18	1.76	0.042
Jul	300-330	0.18	1.76	0.032	0.18	1.76	0.032	0.18	1.76	0.032
Jul	330-360	0.18	1.76	0.055	0.18	1.76	0.055	0.18	1.76	0.055
Aug	0-30	0.18	1.76	0.045	0.18	0.44	0.045	0.18	1.76	0.045
Aug	30-60	0.18	1.76	0.045	0.18	0.44	0.045	0.18	1.76	0.045
Aug	60-90	0.18	1.76	0.056	0.18	0.44	0.056	0.18	1.76	0.056
Aug	90-120	0.18	1.76	0.046	0.18	0.44	0.046	0.18	1.76	0.046
Aug	120-150	0.18	1.76	0.048	0.18	0.44	0.048	0.18	1.76	0.048
Aug	150-180	0.18	1.76	0.063	0.18	0.44	0.063	0.18	1.76	0.063
Aug	180-210	0.18	1.76	0.051	0.18	0.44	0.051	0.18	1.76	0.051
Aug	210-240	0.18	1.76	0.040	0.18	0.44	0.040	0.18	1.76	0.040
Aug	240-270	0.18	1.76	0.043	0.18	0.44	0.043	0.18	1.76	0.043
Aug	270-300	0.18	1.76	0.042	0.18	0.44	0.042	0.18	1.76	0.042
Aug	300-330	0.18	1.76	0.032	0.18	0.44	0.032	0.18	1.76	0.032
Aug	330-360	0.18	1.76	0.055	0.18	0.44	0.055	0.18	1.76	0.055
Sep	0-30	0.18	0.52	0.040	0.18	0.52	0.040	0.18	0.52	0.040
Sep	30-60	0.18	0.52	0.040	0.18	0.52	0.040	0.18	0.52	0.040
Sep	60-90	0.18	0.52	0.053	0.18	0.52	0.053	0.18	0.52	0.053
Sep	90-120	0.18	0.52	0.041	0.18	0.52	0.041	0.18	0.52	0.041
Sep	120-150	0.18	0.52	0.043	0.18	0.52	0.043	0.18	0.52	0.043
Sep	150-180	0.18	0.52	0.059	0.18	0.52	0.059	0.18	0.52	0.059
Sep	180-210	0.18	0.52	0.046	0.18	0.52	0.046	0.18	0.52	0.046
Sep	210-240	0.18	0.52	0.033	0.18	0.52	0.033	0.18	0.52	0.033
Sep	240-270	0.18	0.52	0.037	0.18	0.52	0.037	0.18	0.52	0.037
Sep	270-300	0.18	0.52	0.036	0.18	0.52	0.036	0.18	0.52	0.036
Sep	300-330	0.18	0.52	0.027	0.18	0.52	0.027	0.18	0.52	0.027
Sep	330-360	0.18	0.52	0.051	0.18	0.52	0.051	0.18	0.52	0.051
Oct	0-30	0.18	0.52	0.040	0.18	0.52	0.040	0.18	0.52	0.040
Oct	30-60	0.18	0.52	0.040	0.18	0.52	0.040	0.18	0.52	0.040
Oct	60-90	0.18	0.52	0.053	0.18	0.52	0.053	0.18	0.52	0.053
Oct	90-120	0.18	0.52	0.041	0.18	0.52	0.041	0.18	0.52	0.041
Oct	120-150	0.18	0.52	0.043	0.18	0.52	0.043	0.18	0.52	0.043
Oct	150-180	0.18	0.52	0.059	0.18	0.52	0.059	0.18	0.52	0.059
Oct	180-210	0.18	0.52	0.046	0.18	0.52	0.046	0.18	0.52	0.046
Oct	210-240	0.18	0.52	0.033	0.18	0.52	0.033	0.18	0.52	0.033

Station = IND		2011			2012			2013		
Month	Sector (degrees)	Albedo	Bowen Ratio	Roughness (m)	Albedo	Bowen Ratio	Roughness (m)	Albedo	Bowen Ratio	Roughness (m)
Oct	240-270	0.18	0.52	0.037	0.18	0.52	0.037	0.18	0.52	0.037
Oct	270-300	0.18	0.52	0.036	0.18	0.52	0.036	0.18	0.52	0.036
Oct	300-330	0.18	0.52	0.027	0.18	0.52	0.027	0.18	0.52	0.027
Oct	330-360	0.18	0.52	0.051	0.18	0.52	0.051	0.18	0.52	0.051
Nov	0-30	0.18	0.52	0.040	0.18	2.26	0.040	0.18	0.52	0.040
Nov	30-60	0.18	0.52	0.040	0.18	2.26	0.040	0.18	0.52	0.040
Nov	60-90	0.18	0.52	0.053	0.18	2.26	0.053	0.18	0.52	0.053
Nov	90-120	0.18	0.52	0.041	0.18	2.26	0.041	0.18	0.52	0.041
Nov	120-150	0.18	0.52	0.043	0.18	2.26	0.043	0.18	0.52	0.043
Nov	150-180	0.18	0.52	0.059	0.18	2.26	0.059	0.18	0.52	0.059
Nov	180-210	0.18	0.52	0.046	0.18	2.26	0.046	0.18	0.52	0.046
Nov	210-240	0.18	0.52	0.033	0.18	2.26	0.033	0.18	0.52	0.033
Nov	240-270	0.18	0.52	0.037	0.18	2.26	0.037	0.18	0.52	0.037
Nov	270-300	0.18	0.52	0.036	0.18	2.26	0.036	0.18	0.52	0.036
Nov	300-330	0.18	0.52	0.027	0.18	2.26	0.027	0.18	0.52	0.027
Nov	330-360	0.18	0.52	0.051	0.18	2.26	0.051	0.18	0.52	0.051
Dec	0-30	0.18	0.52	0.032	0.18	0.89	0.032	0.18	0.52	0.032
Dec	30-60	0.18	0.52	0.033	0.18	0.89	0.033	0.18	0.52	0.033
Dec	60-90	0.18	0.52	0.046	0.18	0.89	0.046	0.18	0.52	0.046
Dec	90-120	0.18	0.52	0.030	0.18	0.89	0.030	0.18	0.52	0.030
Dec	120-150	0.18	0.52	0.031	0.18	0.89	0.031	0.18	0.52	0.031
Dec	150-180	0.18	0.52	0.040	0.18	0.89	0.040	0.18	0.52	0.040
Dec	180-210	0.18	0.52	0.027	0.18	0.89	0.027	0.18	0.52	0.027
Dec	210-240	0.18	0.52	0.016	0.18	0.89	0.016	0.18	0.52	0.016
Dec	240-270	0.18	0.52	0.022	0.18	0.89	0.022	0.18	0.52	0.022
Dec	270-300	0.18	0.52	0.022	0.18	0.89	0.022	0.18	0.52	0.022
Dec	300-330	0.18	0.52	0.019	0.18	0.89	0.019	0.18	0.52	0.019
Dec	330-360	0.18	0.52	0.041	0.18	0.89	0.041	0.18	0.52	0.041

Table A-8. Surface characteristics for Tulsa R. L. Jones Jr. (RVS) by month and year.

Station = RVS		2011			2012			2013		
Month	Sector (degrees)	Albedo	Bowen Ratio	Roughness (m)	Albedo	Bowen Ratio	Roughness (m)	Albedo	Bowen Ratio	Roughness (m)
Jan	0-30	0.18	0.48	0.055	0.18	0.87	0.055	0.18	1.96	0.055
Jan	30-60	0.18	0.48	0.031	0.18	0.87	0.031	0.18	1.96	0.031
Jan	60-90	0.18	0.48	0.043	0.18	0.87	0.043	0.18	1.96	0.043
Jan	90-120	0.18	0.48	0.039	0.18	0.87	0.039	0.18	1.96	0.039
Jan	120-150	0.18	0.48	0.030	0.18	0.87	0.030	0.18	1.96	0.030
Jan	150-180	0.18	0.48	0.059	0.18	0.87	0.059	0.18	1.96	0.059
Jan	180-210	0.18	0.48	0.048	0.18	0.87	0.048	0.18	1.96	0.048
Jan	210-240	0.18	0.48	0.110	0.18	0.87	0.110	0.18	1.96	0.110
Jan	240-270	0.18	0.48	0.083	0.18	0.87	0.083	0.18	1.96	0.083
Jan	270-300	0.18	0.48	0.057	0.18	0.87	0.057	0.18	1.96	0.057
Jan	300-330	0.18	0.48	0.083	0.18	0.87	0.083	0.18	1.96	0.083
Jan	330-360	0.18	0.48	0.044	0.18	0.87	0.044	0.18	1.96	0.044
Feb	0-30	0.18	0.48	0.055	0.18	0.87	0.055	0.18	1.96	0.055
Feb	30-60	0.18	0.48	0.031	0.18	0.87	0.031	0.18	1.96	0.031
Feb	60-90	0.18	0.48	0.043	0.18	0.87	0.043	0.18	1.96	0.043
Feb	90-120	0.18	0.48	0.039	0.18	0.87	0.039	0.18	1.96	0.039
Feb	120-150	0.18	0.48	0.030	0.18	0.87	0.030	0.18	1.96	0.030
Feb	150-180	0.18	0.48	0.059	0.18	0.87	0.059	0.18	1.96	0.059
Feb	180-210	0.18	0.48	0.048	0.18	0.87	0.048	0.18	1.96	0.048
Feb	210-240	0.18	0.48	0.110	0.18	0.87	0.110	0.18	1.96	0.110
Feb	240-270	0.18	0.48	0.083	0.18	0.87	0.083	0.18	1.96	0.083
Feb	270-300	0.18	0.48	0.057	0.18	0.87	0.057	0.18	1.96	0.057
Feb	300-330	0.18	0.48	0.083	0.18	0.87	0.083	0.18	1.96	0.083
Feb	330-360	0.18	0.48	0.044	0.18	0.87	0.044	0.18	1.96	0.044
Mar	0-30	0.16	0.36	0.088	0.16	0.56	0.088	0.16	1.37	0.088
Mar	30-60	0.16	0.36	0.056	0.16	0.56	0.056	0.16	1.37	0.056
Mar	60-90	0.16	0.36	0.070	0.16	0.56	0.070	0.16	1.37	0.070
Mar	90-120	0.16	0.36	0.072	0.16	0.56	0.072	0.16	1.37	0.072
Mar	120-150	0.16	0.36	0.052	0.16	0.56	0.052	0.16	1.37	0.052
Mar	150-180	0.16	0.36	0.068	0.16	0.56	0.068	0.16	1.37	0.068
Mar	180-210	0.16	0.36	0.063	0.16	0.56	0.063	0.16	1.37	0.063
Mar	210-240	0.16	0.36	0.219	0.16	0.56	0.219	0.16	1.37	0.219
Mar	240-270	0.16	0.36	0.133	0.16	0.56	0.133	0.16	1.37	0.133
Mar	270-300	0.16	0.36	0.095	0.16	0.56	0.095	0.16	1.37	0.095
Mar	300-330	0.16	0.36	0.133	0.16	0.56	0.133	0.16	1.37	0.133
Mar	330-360	0.16	0.36	0.083	0.16	0.56	0.083	0.16	1.37	0.083
Apr	0-30	0.16	0.36	0.088	0.16	0.56	0.088	0.16	1.37	0.088
Apr	30-60	0.16	0.36	0.056	0.16	0.56	0.056	0.16	1.37	0.056

Station = RVS		2011			2012			2013		
Month	Sector (degrees)	Albedo	Bowen Ratio	Roughness (m)	Albedo	Bowen Ratio	Roughness (m)	Albedo	Bowen Ratio	Roughness (m)
Apr	60-90	0.16	0.36	0.070	0.16	0.56	0.070	0.16	1.37	0.070
Apr	90-120	0.16	0.36	0.072	0.16	0.56	0.072	0.16	1.37	0.072
Apr	120-150	0.16	0.36	0.052	0.16	0.56	0.052	0.16	1.37	0.052
Apr	150-180	0.16	0.36	0.068	0.16	0.56	0.068	0.16	1.37	0.068
Apr	180-210	0.16	0.36	0.063	0.16	0.56	0.063	0.16	1.37	0.063
Apr	210-240	0.16	0.36	0.219	0.16	0.56	0.219	0.16	1.37	0.219
Apr	240-270	0.16	0.36	0.133	0.16	0.56	0.133	0.16	1.37	0.133
Apr	270-300	0.16	0.36	0.095	0.16	0.56	0.095	0.16	1.37	0.095
Apr	300-330	0.16	0.36	0.133	0.16	0.56	0.133	0.16	1.37	0.133
Apr	330-360	0.16	0.36	0.083	0.16	0.56	0.083	0.16	1.37	0.083
May	0-30	0.16	0.36	0.088	0.16	0.56	0.088	0.16	1.37	0.088
May	30-60	0.16	0.36	0.056	0.16	0.56	0.056	0.16	1.37	0.056
May	60-90	0.16	0.36	0.070	0.16	0.56	0.070	0.16	1.37	0.070
May	90-120	0.16	0.36	0.072	0.16	0.56	0.072	0.16	1.37	0.072
May	120-150	0.16	0.36	0.052	0.16	0.56	0.052	0.16	1.37	0.052
May	150-180	0.16	0.36	0.068	0.16	0.56	0.068	0.16	1.37	0.068
May	180-210	0.16	0.36	0.063	0.16	0.56	0.063	0.16	1.37	0.063
May	210-240	0.16	0.36	0.219	0.16	0.56	0.219	0.16	1.37	0.219
May	240-270	0.16	0.36	0.133	0.16	0.56	0.133	0.16	1.37	0.133
May	270-300	0.16	0.36	0.095	0.16	0.56	0.095	0.16	1.37	0.095
May	300-330	0.16	0.36	0.133	0.16	0.56	0.133	0.16	1.37	0.133
May	330-360	0.16	0.36	0.083	0.16	0.56	0.083	0.16	1.37	0.083
Jun	0-30	0.17	0.38	0.250	0.17	0.57	0.250	0.17	1.33	0.250
Jun	30-60	0.17	0.38	0.114	0.17	0.57	0.114	0.17	1.33	0.114
Jun	60-90	0.17	0.38	0.131	0.17	0.57	0.131	0.17	1.33	0.131
Jun	90-120	0.17	0.38	0.138	0.17	0.57	0.138	0.17	1.33	0.138
Jun	120-150	0.17	0.38	0.098	0.17	0.57	0.098	0.17	1.33	0.098
Jun	150-180	0.17	0.38	0.075	0.17	0.57	0.075	0.17	1.33	0.075
Jun	180-210	0.17	0.38	0.107	0.17	0.57	0.107	0.17	1.33	0.107
Jun	210-240	0.17	0.38	0.389	0.17	0.57	0.389	0.17	1.33	0.389
Jun	240-270	0.17	0.38	0.318	0.17	0.57	0.318	0.17	1.33	0.318
Jun	270-300	0.17	0.38	0.265	0.17	0.57	0.265	0.17	1.33	0.265
Jun	300-330	0.17	0.38	0.325	0.17	0.57	0.325	0.17	1.33	0.325
Jun	330-360	0.17	0.38	0.244	0.17	0.57	0.244	0.17	1.33	0.244
Jul	0-30	0.17	0.38	0.250	0.17	0.57	0.250	0.17	1.33	0.250
Jul	30-60	0.17	0.38	0.114	0.17	0.57	0.114	0.17	1.33	0.114
Jul	60-90	0.17	0.38	0.131	0.17	0.57	0.131	0.17	1.33	0.131
Jul	90-120	0.17	0.38	0.138	0.17	0.57	0.138	0.17	1.33	0.138
Jul	120-150	0.17	0.38	0.098	0.17	0.57	0.098	0.17	1.33	0.098

Station = RVS		2011			2012			2013		
Month	Sector (degrees)	Albedo	Bowen Ratio	Roughness (m)	Albedo	Bowen Ratio	Roughness (m)	Albedo	Bowen Ratio	Roughness (m)
Jul	150-180	0.17	0.38	0.075	0.17	0.57	0.075	0.17	1.33	0.075
Jul	180-210	0.17	0.38	0.107	0.17	0.57	0.107	0.17	1.33	0.107
Jul	210-240	0.17	0.38	0.389	0.17	0.57	0.389	0.17	1.33	0.389
Jul	240-270	0.17	0.38	0.318	0.17	0.57	0.318	0.17	1.33	0.318
Jul	270-300	0.17	0.38	0.265	0.17	0.57	0.265	0.17	1.33	0.265
Jul	300-330	0.17	0.38	0.325	0.17	0.57	0.325	0.17	1.33	0.325
Jul	330-360	0.17	0.38	0.244	0.17	0.57	0.244	0.17	1.33	0.244
Aug	0-30	0.17	0.38	0.250	0.17	0.57	0.250	0.17	1.33	0.250
Aug	30-60	0.17	0.38	0.114	0.17	0.57	0.114	0.17	1.33	0.114
Aug	60-90	0.17	0.38	0.131	0.17	0.57	0.131	0.17	1.33	0.131
Aug	90-120	0.17	0.38	0.138	0.17	0.57	0.138	0.17	1.33	0.138
Aug	120-150	0.17	0.38	0.098	0.17	0.57	0.098	0.17	1.33	0.098
Aug	150-180	0.17	0.38	0.075	0.17	0.57	0.075	0.17	1.33	0.075
Aug	180-210	0.17	0.38	0.107	0.17	0.57	0.107	0.17	1.33	0.107
Aug	210-240	0.17	0.38	0.389	0.17	0.57	0.389	0.17	1.33	0.389
Aug	240-270	0.17	0.38	0.318	0.17	0.57	0.318	0.17	1.33	0.318
Aug	270-300	0.17	0.38	0.265	0.17	0.57	0.265	0.17	1.33	0.265
Aug	300-330	0.17	0.38	0.325	0.17	0.57	0.325	0.17	1.33	0.325
Aug	330-360	0.17	0.38	0.244	0.17	0.57	0.244	0.17	1.33	0.244
Sep	0-30	0.17	0.48	0.250	0.17	0.87	0.250	0.17	1.96	0.250
Sep	30-60	0.17	0.48	0.114	0.17	0.87	0.114	0.17	1.96	0.114
Sep	60-90	0.17	0.48	0.131	0.17	0.87	0.131	0.17	1.96	0.131
Sep	90-120	0.17	0.48	0.138	0.17	0.87	0.138	0.17	1.96	0.138
Sep	120-150	0.17	0.48	0.098	0.17	0.87	0.098	0.17	1.96	0.098
Sep	150-180	0.17	0.48	0.075	0.17	0.87	0.075	0.17	1.96	0.075
Sep	180-210	0.17	0.48	0.107	0.17	0.87	0.107	0.17	1.96	0.107
Sep	210-240	0.17	0.48	0.389	0.17	0.87	0.389	0.17	1.96	0.389
Sep	240-270	0.17	0.48	0.318	0.17	0.87	0.318	0.17	1.96	0.318
Sep	270-300	0.17	0.48	0.265	0.17	0.87	0.265	0.17	1.96	0.265
Sep	300-330	0.17	0.48	0.325	0.17	0.87	0.325	0.17	1.96	0.325
Sep	330-360	0.17	0.48	0.244	0.17	0.87	0.244	0.17	1.96	0.244
Oct	0-30	0.17	0.48	0.250	0.17	0.87	0.250	0.17	1.96	0.250
Oct	30-60	0.17	0.48	0.114	0.17	0.87	0.114	0.17	1.96	0.114
Oct	60-90	0.17	0.48	0.131	0.17	0.87	0.131	0.17	1.96	0.131
Oct	90-120	0.17	0.48	0.138	0.17	0.87	0.138	0.17	1.96	0.138
Oct	120-150	0.17	0.48	0.098	0.17	0.87	0.098	0.17	1.96	0.098
Oct	150-180	0.17	0.48	0.075	0.17	0.87	0.075	0.17	1.96	0.075
Oct	180-210	0.17	0.48	0.107	0.17	0.87	0.107	0.17	1.96	0.107
Oct	210-240	0.17	0.48	0.389	0.17	0.87	0.389	0.17	1.96	0.389

Station = RVS		2011			2012			2013		
Month	Sector (degrees)	Albedo	Bowen Ratio	Roughness (m)	Albedo	Bowen Ratio	Roughness (m)	Albedo	Bowen Ratio	Roughness (m)
Oct	240-270	0.17	0.48	0.318	0.17	0.87	0.318	0.17	1.96	0.318
Oct	270-300	0.17	0.48	0.265	0.17	0.87	0.265	0.17	1.96	0.265
Oct	300-330	0.17	0.48	0.325	0.17	0.87	0.325	0.17	1.96	0.325
Oct	330-360	0.17	0.48	0.244	0.17	0.87	0.244	0.17	1.96	0.244
Nov	0-30	0.17	0.48	0.250	0.17	0.87	0.250	0.17	1.96	0.250
Nov	30-60	0.17	0.48	0.114	0.17	0.87	0.114	0.17	1.96	0.114
Nov	60-90	0.17	0.48	0.131	0.17	0.87	0.131	0.17	1.96	0.131
Nov	90-120	0.17	0.48	0.138	0.17	0.87	0.138	0.17	1.96	0.138
Nov	120-150	0.17	0.48	0.098	0.17	0.87	0.098	0.17	1.96	0.098
Nov	150-180	0.17	0.48	0.075	0.17	0.87	0.075	0.17	1.96	0.075
Nov	180-210	0.17	0.48	0.107	0.17	0.87	0.107	0.17	1.96	0.107
Nov	210-240	0.17	0.48	0.389	0.17	0.87	0.389	0.17	1.96	0.389
Nov	240-270	0.17	0.48	0.318	0.17	0.87	0.318	0.17	1.96	0.318
Nov	270-300	0.17	0.48	0.265	0.17	0.87	0.265	0.17	1.96	0.265
Nov	300-330	0.17	0.48	0.325	0.17	0.87	0.325	0.17	1.96	0.325
Nov	330-360	0.17	0.48	0.244	0.17	0.87	0.244	0.17	1.96	0.244
Dec	0-30	0.18	0.48	0.055	0.18	0.87	0.055	0.18	1.96	0.055
Dec	30-60	0.18	0.48	0.031	0.18	0.87	0.031	0.18	1.96	0.031
Dec	60-90	0.18	0.48	0.043	0.18	0.87	0.043	0.18	1.96	0.043
Dec	90-120	0.18	0.48	0.039	0.18	0.87	0.039	0.18	1.96	0.039
Dec	120-150	0.18	0.48	0.030	0.18	0.87	0.030	0.18	1.96	0.030
Dec	150-180	0.18	0.48	0.059	0.18	0.87	0.059	0.18	1.96	0.059
Dec	180-210	0.18	0.48	0.048	0.18	0.87	0.048	0.18	1.96	0.048
Dec	210-240	0.18	0.48	0.110	0.18	0.87	0.110	0.18	1.96	0.110
Dec	240-270	0.18	0.48	0.083	0.18	0.87	0.083	0.18	1.96	0.083
Dec	270-300	0.18	0.48	0.057	0.18	0.87	0.057	0.18	1.96	0.057
Dec	300-330	0.18	0.48	0.083	0.18	0.87	0.083	0.18	1.96	0.083
Dec	330-360	0.18	0.48	0.044	0.18	0.87	0.044	0.18	1.96	0.044

REFERENCES

U.S. EPA. (2013). AERSURFACE User’s Guide. U.S. Environmental Protection Agency. EPA 454/B-08-001. Revised January 16, 2013.

U.S. EPA. (2015). AERMINUTE User’s Guide. U.S. Environmental Protection Agency. EPA 454/B-15-006.

U.S. EPA. (2016a). User's Guide for the AMS/EPA Regulatory Model – AERMOD. U.S. Environmental Protection Agency. 454/B-16-011.

U.S. EPA. (2016b). User’s Guide for the AERMOD Meteorological Processor (AERMET). U.S. Environmental Protection Agency. EPA-454/B-16-010.

APPENDIX B

DEVELOPMENT OF HOURLY EMISSIONS PROFILES

Preface: The source type influenced how the hourly emissions profiles were developed. The methods followed are summarized below separately for EGU and other sources.

B.1 EGU Sources

The NEI stores references to the Office of Regulatory Information Systems (ORIS) identification code for most sources that have Continuous Emissions Monitoring System (CEMS) data in the CAMD database. For these stacks the relative hourly profiles were derived from the hourly values in the CAMD database, and the annual emissions totals were taken from the NEI (Table B-1). EGU emissions came from the NEI for their respective years. Where CEMS data was available, the CEMS emissions values were used and the emissions in the annual inventory were adjusted to match the temporal pattern of the year-specific CEMS data. The EGU units with more than 20 tons of SO_2 emissions in at least one year for which CEMS data are available are listed in Table B-1 along with their annual SO_2 emissions for 2011, 2012, and 2013. Sources at the SEMASS Partnership facility (county 25023 and facility ID 8127611) and IP&L – Harding Street (county 18097 and facility ID 7255211) are designated as EGUs but are not matched to sources in the CAMD database. These sources were temporalized to hourly values using average temporal profiles that were derived based on other EGU units in their respective regions.

Table B-1. SO_2 emissions each year for EGUs included in the air quality modeling.

FIPS	Facility Name	Facility ID	Unit ID	2011	2012	2013
25005	BRAYTON POINT ENERGY LLC	5058411	87339613	3,535	1,228	1,625
25005	BRAYTON POINT ENERGY LLC	5058411	87339713	45	12	118
25005	BRAYTON POINT ENERGY LLC	5058411	87340713	4,298	1,859	1,383
25005	BRAYTON POINT ENERGY LLC	5058411	87340813	10,769	6,033	4,479
18097	IP&L - HARDING STREET	7255211	91188613	8,634	10,531	13,324
18097	IP&L - HARDING STREET	7255211	91188713	7,941	10,270	12,603
18097	IP&L - HARDING STREET	7255211	91188813	681	632	1,846
18097	IP&L - HARDING STREET	7255211	91188813	1,739	109	200
40131	PSO NORTHEASTERN PWR STA	8212411	6698813		8,039	9,008
40131	PSO NORTHEASTERN PWR STA	8212411	6698813	8,879		
40131	PSO NORTHEASTERN PWR STA	8212411	6698813		20	38
40131	PSO NORTHEASTERN PWR STA	8212411	6698813	26		
40131	PSO NORTHEASTERN PWR STA	8212411	6698313		7,402	9,337
40131	PSO NORTHEASTERN PWR STA	8212411	6698313	9,008		
40131	PSO NORTHEASTERN PWR STA	8212411	6698313		27	22
40131	PSO NORTHEASTERN PWR STA	8212411	6698313	26		

B.2 Non-EGU Sources

For non-EGU sources that did not have hourly SO_2 data in the CAMD database, SCC-specific temporal profiles from EPA's 2011v6.3 emissions modeling platform were used to prepare the hourly factors. Stacks with emissions greater than 20 tons of SO_2 in 2011, 2012, or 2013 for which temporal profiles were used are listed in Table B-2 below. The allocation of the sources to the hourly factors needed for AERMOD was done using tools available within the Sparse Matrix Operator Kernel Emissions (SMOKE) modeling system version 4.5 (UNC, 2017). The tools support the generation of "helper files" from which the AERMOD input files can be derived. The temporal values output from SMOKE were renormalized from scalars to factors that sum to 1 to aid with quality assurance and usability of the factors.

Table B-2. SO_2 emissions each year for non-EGU release points included in the air quality modeling[1].

FIPS	Facility Name	Facility ID	Unit ID	Release	2011	2012	2013
18097	Citizens Thermal	4885311	100805413	30985212	2,094	1,849	1,575
18097	Citizens Thermal	4885311	100805713	30985012	1,029	853	855
18097	Citizens Thermal	4885311	100805813	30985012	1,225	1,150	1,375
40037	SAPULPA	7320611	72251213	66374812	79	79	98
40037	SAPULPA	7320611	8331413	8217312	33	33	34
40037	SAPULPA	7320611	8331213	8217212	100	100	108
18097	VERTELLUS AGRICULTURE &	7972111	65408713	60023412	20	17	20
18097	QUEMETCO, INC.	8235411	65358713	5022512	49	49	16
18097	QUEMETCO, INC.	8235411	65358713	5022612	71	71	69
40143	TULSA RFNRY WEST	8402711	654613	655312	103	42	26
40143	TULSA RFNRY WEST	8402711	654413	660012	45	20	9
40143	TULSA RFNRY WEST	8402711	654313	659912	380	237	169
40143	TULSA RFNRY WEST	8402711	654113	663512	36	18	11
40143	TULSA RFNRY WEST	8402711	651713	655012	59	65	24
40143	TULSA RFNRY WEST	8402711	651413	661212	270	210	125
40143	TULSA RFNRY WEST	8402711	651313	658812	43	41	11
40143	TULSA RFNRY WEST	8402711	651113	662812	39		
40143	TULSA RFNRY WEST	8402711	651113	662812		43	17
40143	TULSA RFNRY WEST	8402711	651013	658912	157	150	37
40143	TULSA RFNRY WEST	8402711	650913	654912	74	55	34
40143	TULSA RFNRY WEST	8402711	650813	656012	38	46	8
40143	TULSA RFNRY WEST	8402711	663113	651512	866	688	360
40143	TULSA RFNRY WEST	8402711	658713	651412	460	370	211

The emissions factors developed for non-EGU sources were monthly, hour-of-day, or month-hour-of-day, where day was weekday, Saturday, or Sunday. These emission factors correspond to the MONTH, HROFDY, and MHRDOW emission factors used in AERMOD (U.S. EPA, 2016). These emission factors are set to sum to 1 for each source. For example, for a source using the MONTH emission factors, the 12 monthly factors sum to 1. This means that a particular month's factor allocates a portion of the annual emissions to that month. Further processing is needed to create hourly emissions for the sources. For monthly factors, the monthly factor is divided by the number of hours in the month (number of days x 24 hours) and this ratio is multiplied by the annual emissions to get an hourly emission rate and this rate is then converted to a g/s rate. This rate is then input into AERMOD as the MONTH emission factor, and the reference emission rate in AERMOD (emission rate on the SRCPARAM line in the

[1] Based on units emitting over 20 tons of SO_2.

AERMOD input file) is set to 1.0. This method creates an hourly emission rate while conserving the annual emissions.

Consider a source with the following monthly factors (Table B-3) output from SMOKE for 2011 and annual emissions of 100.32 tons. The factors divide the emissions equally across the months, resulting in the monthly emissions (in tons) shown for each month. To convert the monthly emissions for a given month, to g/s, the following equation is used:

$$E_{hour} = E_{annual} \times \left(\frac{1}{Days_{month}} \right) \times \left(\frac{1}{24} \right) \times 251.9957778 \qquad \text{Equation B-1}$$

Where E_{hour} is the hourly emission rate in g/s, E_{annual} are the annual emissions in tons, $Days_{month}$ are the number of days in the month (31 days for January, etc.), 1/24 is the reciprocal of the number of hours in a day, and 251.9957778 is the conversion factor to convert from tons/hour to g/s. The resulting hourly emissions rates are also shown in Table B-3. Figure B-1 shows how the hourly emissions are input into AERMOD using the SRCPARAM and EMISFACT keywords. Equation 1 is also used to calculate the MHRDOW emissions and a similar form of Equation 1 is used for HROFDY emissions, with the exception that $1/Days_{month}$ is 1/365 (number of days in the year).

Table B-3. Example calculation of hourly emissions using the SMOKE MONTH temporal factors for 2011.

Month	SMOKE factor	$Days_{month}$	E_{hour} (g/s)
January	0.083333	31	2.831565
February	0.083333	28	3.134947
March	0.083333	31	2.831565
April	0.083333	30	2.925951
May	0.083333	31	2.831565
June	0.083333	30	2.925951
July	0.083333	31	2.831565
August	0.083333	31	2.831565
September	0.083333	30	2.925951
October	0.083333	31	2.831565
November	0.083333	30	2.925951
December	0.083333	31	2.831565

```
SO SRCPARAM SAP_SN1      1.000000E+00   28.35000  530.37000    9.60000    1.86000
SO EMISFACT SAP_SN1     MONTH     2.831565E+00  3.134947E+00
SO EMISFACT SAP_SN1     MONTH     2.831565E+00  2.925951E+00
SO EMISFACT SAP_SN1     MONTH     2.831565E+00  2.925951E+00
SO EMISFACT SAP_SN1     MONTH     2.831565E+00  2.831565E+00
SO EMISFACT SAP_SN1     MONTH     2.925951E+00  2.831565E+00
SO EMISFACT SAP_SN1     MONTH     2.925951E+00  2.831565E+00
```

Figure B-1. Example AERMOD input emission lines for monthly emissions.

B.3 AERMOD inputs

Tables B-4 through B-41 list the cross walks between facility unit identifiers and AERMOD source identifiers and the 2011-2013 AERMOD inputs for each of the three study areas. Note that the AERMOD source identifiers are unique to each year. In some cases, a particular emission release point may not have an AERMOD source identifier for one year but may have an identifier for other years. Years in which a release point does not have an AERMOD identifier are left as blanks.

Table B-4. Fall River 2011-2013 AERMOD source identifier crosswalk.

Facility Name	Unit ID	Process ID	Release Point ID	AERMOD 2011	AERMOD 2012	AERMOD 2013
BRAYTON POINT ENERGY LLC	87339613	83612912	118371314	BRAY_SE1	BRAY_SE1	BRAY_SE1
BRAYTON POINT ENERGY LLC	87339613	83612912	118371714	BRAY_SE2	BRAY_SE2	BRAY_SE2
BRAYTON POINT ENERGY LLC	87339713	83613312	118371814	BRAY_SE3	BRAY_SE3	BRAY_SE3
BRAYTON POINT ENERGY LLC	87339713	83613312	118371914	BRAY_SE4	BRAY_SE4	BRAY_SE4
BRAYTON POINT ENERGY LLC	87339713	83613312	118372014		BRAY_SE5	
BRAYTON POINT ENERGY LLC	87339713	83613312	118372114	BRAY_SE5	BRAY_SE6	BRAY_SE5
BRAYTON POINT ENERGY LLC	87340713	83612812	118373214	BRAY_SE6	BRAY_SE7	BRAY_SE6
BRAYTON POINT ENERGY LLC	87340713	83612812	118373514	BRAY_SE7	BRAY_SE8	BRAY_SE7
BRAYTON POINT ENERGY LLC	87340813	83612612	118373614	BRAY_SE8	BRAY_SE9	BRAY_SE8
BRAYTON POINT ENERGY LLC	87340813	83612612	118373714	BRAY_SE9	BRAY_SE10	BRAY_SE9
BRAYTON POINT ENERGY LLC	90543213	83613612	122762214	BRAY_SN1	BRAY_SN1	BRAY_SN1
BRAYTON POINT ENERGY LLC	90543413	83613612	122762414	BRAY_SN1	BRAY_SN1	BRAY_SN1
BRAYTON POINT ENERGY LLC	87341513	83613212	118374814	BRAY_SN2	BRAY_SN2	BRAY_SN2
BRAYTON POINT ENERGY LLC	87341613	83612512	118374914	BRAY_SN2	BRAY_SN2	BRAY_SN2

Table B-5. 2011 Fall River point source emissions, locations, and stack parameters.

Facility Name	AERMDOD source ID	Emissions (tons year-1)	Emission factor	UTM-x (m)	UTM-y (m)	Elevation (m)	Stack height (m)	Stack temperature (K)	Stack velocity (m s-1)	Stack diameter (m)
BRAYTON POINT ENERGY LLC	BRAY_SE1	3534.90	HOURLY	317613.67	4620047.98	5.07	107.29	383.15	20.45	4.42
BRAYTON POINT ENERGY LLC	BRAY_SE2	0.01	HOURLY	317613.67	4620047.98	5.07	107.29	383.15	20.45	4.42
BRAYTON POINT ENERGY LLC	BRAY_SE3	45.24	HOURLY	317536.89	4620117.91	4.72	152.40	432.04	21.73	5.64
BRAYTON POINT ENERGY LLC	BRAY_SE4	0.77	HOURLY	317536.89	4620117.91	4.72	152.40	432.04	21.73	5.64
BRAYTON POINT ENERGY LLC	BRAY_SE5	0.11	HOURLY	317536.89	4620117.91	4.72	152.40	432.04	21.73	5.64
BRAYTON POINT ENERGY LLC	BRAY_SE6	4298.40	HOURLY	317639.35	4620024.01	5.56	107.29	383.15	20.45	4.42
BRAYTON POINT ENERGY LLC	BRAY_SE7	0.01	HOURLY	317639.35	4620024.01	5.56	107.29	383.15	20.45	4.42
BRAYTON POINT ENERGY LLC	BRAY_SE8	0.02	HOURLY	317577.42	4620064.54	4.81	107.29	405.37	24.93	5.94
BRAYTON POINT ENERGY LLC	BRAY_SE9	10769.00	HOURLY	317577.42	4620064.54	4.81	107.29	405.37	24.93	5.94
BRAYTON POINT ENERGY LLC	BRAY_SN2	0.0004	MONTH	317600.47	4619900.00	8.20	3.66	783.15	24.66	0.30

Table B-6. 2011 Fall River area source emissions, locations, and stack parameters.

Facility Name	AERMDOD source ID	Emissions (tons year-1)	Emission factor	UTM-x (m)	UTM-y (m)	Elevation (m)	Release height (m)	X-dimension (m)	Y-dimensio n (m)	Angle	σ_z (m)
BRAYTON POINT ENERGY LLC	BRAY_SN1	3534.90	MONTH	317600.47	4619900.00	8.20	3.05	10.0	10.0	0.0	0.0

Table B-7. 2012 Fall River point source emissions, locations, and stack parameters.

Facility Name	AERMDOD source ID	Emissions (tons year^{-1})	Emission factor	UTM-x (m)	UTM-y (m)	Elevation (m)	Stack height (m)	Stack temperature (K)	Stack velocity (m s^{-1})	Stack diameter (m)
BRAYTON POINT ENERGY LLC	BRAY_SE1	1228.40	HOURLY	317613.67	4620047.98	5.07	107.29	383.15	20.45	4.42
BRAYTON POINT ENERGY LLC	BRAY_SE2	0.08	HOURLY	317613.67	4620047.98	5.07	107.29	383.15	20.45	4.42
BRAYTON POINT ENERGY LLC	BRAY_SE3	12.14	HOURLY	317536.89	4620117.91	4.72	152.40	432.04	21.73	5.64
BRAYTON POINT ENERGY LLC	BRAY_SE4	1.81	HOURLY	317536.89	4620117.91	4.72	152.40	432.04	21.73	5.64
BRAYTON POINT ENERGY LLC	BRAY_SE5	1.53	HOURLY	317536.89	4620117.91	4.72	152.40	432.04	21.73	5.64
BRAYTON POINT ENERGY LLC	BRAY_SE6	0.33	HOURLY	317639.35	4620024.01	5.56	107.29	383.15	20.45	4.42
BRAYTON POINT ENERGY LLC	BRAY_SE7	1859.40	HOURLY	317639.35	4620024.01	5.56	107.29	383.15	20.45	4.42
BRAYTON POINT ENERGY LLC	BRAY_SE8	0.17	HOURLY	317577.42	4620064.54	4.81	107.29	405.37	24.93	5.94
BRAYTON POINT ENERGY LLC	BRAY_SE9	0.13	HOURLY	317577.42	4620064.54	4.81	107.29	405.37	24.93	5.94
BRAYTON POINT ENERGY LLC	BRAY_SE10	6033.0	HOURLY	317577.42	4620064.54	4.81	107.29	405.37	24.93	5.94
BRAYTON POINT ENERGY LLC	BRAY_SN2	0.0014	MONTH	317600.47	4619900.00	8.20	3.66	783.15	24.66	0.30

Table B-8. 2012 Fall River area source emissions, locations, and stack parameters.

Facility Name	AERMDOD source ID	Emissions (tons year^{-1})	Emission factor	UTM-x (m)	UTM-y (m)	Elevation (m)	Release height (m)	X-dimension (m)	Y-dimension (m)	Angle	σ_z (m)
BRAYTON POINT ENERGY LLC	BRAY_SN1	0.008	MONTH	317600.47	4619900.00	8.20	3.05	10.0	10.0	0.0	0.0

Table B-9. 2013 Fall River point source emissions, locations, and stack parameters.

Facility Name	AERMDOD source ID	Emissions (tons year^{-1})	Emission factor	UTM-x (m)	UTM-y (m)	Elevation (m)	Stack height (m)	Stack temperature (K)	Stack velocity (m s^{-1})	Stack diameter (m)
BRAYTON POINT ENERGY LLC	BRAY_SE1	1625.20	HOURLY	317613.67	4620047.98	5.07	107.29	383.15	20.45	4.42
BRAYTON POINT ENERGY LLC	BRAY_SE2	0.01	HOURLY	317613.67	4620047.98	5.07	107.29	383.15	20.45	4.42
BRAYTON POINT ENERGY LLC	BRAY_SE3	118.06	HOURLY	317536.89	4620117.91	4.72	152.40	432.04	21.73	5.64
BRAYTON POINT ENERGY LLC	BRAY_SE4	0.77	HOURLY	317536.89	4620117.91	4.72	152.40	432.04	21.73	5.64
BRAYTON POINT ENERGY LLC	BRAY_SE5	0.11	HOURLY	317536.89	4620117.91	4.72	152.40	432.04	21.73	5.64
BRAYTON POINT ENERGY LLC	BRAY_SE6	1383.00	HOURLY	317639.35	4620024.01	5.56	107.29	383.15	20.45	4.42
BRAYTON POINT ENERGY LLC	BRAY_SE7	0.01	HOURLY	317639.35	4620024.01	5.56	107.29	383.15	20.45	4.42
BRAYTON POINT ENERGY LLC	BRAY_SE8	0.02	HOURLY	317577.42	4620064.54	4.81	107.29	405.37	24.93	5.94
BRAYTON POINT ENERGY LLC	BRAY_SE9	4479.30	HOURLY	317577.42	4620064.54	4.81	107.29	405.37	24.93	5.94
BRAYTON POINT ENERGY LLC	BRAY_SN2	0.0004	MONTH	317600.47	4619900.00	8.20	3.66	783.15	24.66	0.30

Table B-10. 2013 Fall River area source emissions, locations, and stack parameters.

Facility Name	AERMDOD source ID	Emissions (tons year^{-1})	Emission factor	UTM-x (m)	UTM-y (m)	Elevation (m)	Release height (m)	X-dimension (m)	Y-dimension (m)	Angle	σ_z (m)
BRAYTON POINT ENERGY LLC	BRAY_SN1	0.0005	MONTH	317600.47	4619900.00	8.20	3.05	10.0	10.0	0.0	0.0

Table B-11. Indianapolis, IN Indianapolis Belmont WWTP 2011-2013 AERMOD source identifier crosswalk.

Facility Name	Unit ID	Process ID	Release Point ID	AERMOD 2011	AERMOD 2012	AERMOD 2013
INDIANAPOLIS BELMONT WWTP	68272413	64154812	124267014	BELL_SN1	BELL_SN1	
INDIANAPOLIS BELMONT WWTP	68272613	64155012	124267214	BELL_SN1	BELL_SN1	
INDIANAPOLIS BELMONT WWTP	32403713	30985312	123964514	BELL_SN1	BELL_SN1	BELL_SN1
INDIANAPOLIS BELMONT WWTP	32403813	30985312	123964614	BELL_SN1	BELL_SN1	BELL_SN1
INDIANAPOLIS BELMONT WWTP	32403913	30985312	123964814	BELL_SN1	BELL_SN1	BELL_SN1
INDIANAPOLIS BELMONT WWTP	32404013	30985312	123964714	BELL_SN1	BELL_SN1	BELL_SN1

Table B-12. Indianapolis, IN Citizens Thermal 2011-2013 AERMOD source identifier crosswalk.

Facility Name	Unit ID	Process ID	Release Point ID	AERMOD 2011	AERMOD 2012	AERMOD 2013
Citizens Thermal	100805713	30985012	141379114	CIT_SN1	CIT_SN1	CIT_SN1
Citizens Thermal	100805813	30985012	141379414	CIT_SN1	CIT_SN1	CIT_SN1
Citizens Thermal	100805413	30985212	141378314	CIT_SN2	CIT_SN2	CIT_SN2
Citizens Thermal	100805713	30985012	141379214		CIT_SN3	CIT_SN3
Citizens Thermal	100805813	30985012	141379514		CIT_SN3	CIT_SN3
Citizens Thermal	100805313	30984812	141378114	CIT_SN3	CIT_SN4	CIT_SN4
Citizens Thermal	100805413	30984812	141378414	CIT_SN3	CIT_SN4	CIT_SN4
Citizens Thermal	100805513	30985212	141378714	CIT_SN4	CIT_SN5	CIT_SN5
Citizens Thermal	100805613	30985212	141378914	CIT_SN4	CIT_SN5	CIT_SN5
Citizens Thermal	100805913	30984812	141379714	CIT_SN5	CIT_SN6	CIT_SN6
Citizens Thermal	100806013	30984812	141379914	CIT_SN5	CIT_SN6	CIT_SN6

Table B-13. Indianapolis, IN IP&L Harding Street 2011-2013 AERMOD source identifier crosswalk.

Facility Name	Unit ID	Process ID	Release Point ID	AERMOD 2011	AERMOD 2012	AERMOD 2013
IP&L - HARDING STREET	91608313	87281612	124834714	IPL_SE1	IPL_SE1	IPL_SE1
IP&L - HARDING STREET	91608413	87281712	124834814	IPL_SE1	IPL_SE1	IPL_SE1
IP&L - HARDING STREET	91608213	87281512	124834614	IPL_SE2	IPL_SE2	IPL_SE2
IP&L - HARDING STREET	91188213	87281812	123965914	IPL_SE3		IPL_SE3
IP&L - HARDING STREET	91188313	87281912	123966114			IPL_SE4
IP&L - HARDING STREET	91188613	87281212	123966614	IPL_SE4	IPL_SE3	IPL_SE5
IP&L - HARDING STREET	91188713	87281312	123966814	IPL_SE5	IPL_SE4	IPL_SE6
IP&L - HARDING STREET	91188813	87281412	123966914	IPL_SE6	IPL_SE5	IPL_SE7
IP&L - HARDING STREET	91188813	101276612	123967114	IPL_SE7	IPL_SE6	IPL_SE8
IP&L - HARDING STREET	91608513	88573012	124834914	IPL_SE8	IPL_SE7	IPL_SE9

Table B-14. Indianapolis, IN Rolls Royce 2011-2013 AERMOD source identifier crosswalk.

Facility Name	Unit ID	Process ID	Release Point ID	AERMOD 2011	AERMOD 2012	AERMOD 2013
ROLLS ROYCE CORPORATION	68294413	64180912	124304714	RR_SN1	RR_SN1	RR_SN1
ROLLS ROYCE CORPORATION	68294313	64180812	124304614		RR_SN2	
ROLLS ROYCE CORPORATION	2995413	2866112	124164914			RR_SN2
ROLLS ROYCE CORPORATION	2996413	2865012	124166414	RR_SN2	RR_SN3	RR_SN3
ROLLS ROYCE CORPORATION	2996513	2865912	124166514	RR_SN3	RR_SN4	RR_SN4
ROLLS ROYCE CORPORATION	2996613	2865112	124166614	RR_SN3	RR_SN4	RR_SN4
ROLLS ROYCE CORPORATION	2995313	2864812	124164814	RR_SN4	RR_SN5	RR_SN5
ROLLS ROYCE CORPORATION	2995813	64180712	124165414	RR_SN5	RR_SN6	RR_SN6
ROLLS ROYCE CORPORATION	2995413	2866112	124165114	RR_SN6	RR_SN7	RR_SN7
ROLLS ROYCE CORPORATION	2995313	2864812	124164714	RR_SN7		
ROLLS ROYCE CORPORATION	2995413	2866112	124165014	RR_SN8	RR_SN8	RR_SN8
ROLLS ROYCE CORPORATION	2997413	2865812	41165514	RR_SN9	RR_SN9	RR_SN9
ROLLS ROYCE CORPORATION	2994913	2866312	124166214	RR_SN10	RR_SN10	RR_SN10
ROLLS ROYCE CORPORATION	2996113	2866812	124166014	RR_SN10	RR_SN10	RR_SN10
ROLLS ROYCE CORPORATION	2994913	2866312	124166114	RR_SN11	RR_SN11	RR_SN11
ROLLS ROYCE CORPORATION	2996113	2866812	124165914	RR_SN11	RR_SN11	RR_SN11
ROLLS ROYCE CORPORATION	2997513	2865612	124165814	RR_SN12	RR_SN12	RR_SN12
ROLLS ROYCE CORPORATION	2997613	2864712	124165714	RR_SN12	RR_SN12	RR_SN12
ROLLS ROYCE CORPORATION	2995913	2866412	124165614	RR_SN13	RR_SN13	RR_SN13
ROLLS ROYCE CORPORATION	2996213	2865412	124165514	RR_SN13	RR_SN13	RR_SN13
ROLLS ROYCE CORPORATION	2997413	2865812	124167114	RR_SN14	RR_SN14	RR_SN14
ROLLS ROYCE CORPORATION	2996713	2866912	124166714	RR_SN15	RR_SN15	RR_SN15

Table B-15. Indianapolis, IN Vertellus 2011-2013 AERMOD source identifier crosswalk.

Facility Name	Unit ID	Process ID	Release Point ID	AERMOD 2011	AERMOD 2012	AERMOD 2013
VERTELLUS AGRICULTURE & NUTRITION SPECIALTIES LLC	65408713	60023312	90663014	VERT_SN1	VERT_SN1	
VERTELLUS AGRICULTURE & NUTRITION SPECIALTIES LLC	65408713	60023312	141512314	VERT_SN1	VERT_SN1	
VERTELLUS AGRICULTURE & NUTRITION SPECIALTIES LLC	65408713	60023412	90662914	VERT_SN2	VERT_SN2	VERT_SN1
VERTELLUS AGRICULTURE & NUTRITION SPECIALTIES LLC	65408713	60023412	90663314	VERT_SN2	VERT_SN2	VERT_SN1
VERTELLUS AGRICULTURE & NUTRITION SPECIALTIES LLC	65408713	60023412	90663414	VERT_SN2	VERT_SN2	VERT_SN1
VERTELLUS AGRICULTURE & NUTRITION SPECIALTIES LLC	65408713	101303012	90662214	VERT_SN4	VERT_SN4	
VERTELLUS AGRICULTURE & NUTRITION SPECIALTIES LLC	65408613	2863012	90661214	VERT_SN12	VERT_SN12	VERT_SN10
VERTELLUS AGRICULTURE & NUTRITION SPECIALTIES LLC	65408613	2861612	141511014	VERT_SN13	VERT_SN13	VERT_SN11
VERTELLUS AGRICULTURE & NUTRITION SPECIALTIES LLC	65408613	2864112	90660614	VERT_SN14	VERT_SN14	VERT_SN12
VERTELLUS AGRICULTURE & NUTRITION SPECIALTIES LLC	65408613	2863312	90660214	VERT_SN15	VERT_SN15	VERT_SN13

Table B-16. Indianapolis, IN Quemetco 2011-2013 AERMOD source identifier crosswalk.

Facility Name	Unit ID	Process ID	Release Point ID	AERMOD 2011	AERMOD 2012	AERMOD 2013
QUEMETCO, INC.	65358713	5022612	90566814	QUE_SN1	QUE_SN1	QUE_SN1
QUEMETCO, INC.	65358913	5022612	90567014	QUE_SN1	QUE_SN1	QUE_SN1
QUEMETCO, INC.	65358713	5022612	90566714			QUE_SN1
QUEMETCO, INC.	109197013	112719612	154715314			QUE_SN2
QUEMETCO, INC.	65358713	5022512	90566614	QUE_SN2	QUE_SN2	QUE_SN3
QUEMETCO, INC.	65359113	5022512	90567214	QUE_SN2	QUE_SN2	QUE_SN3

Table B-17. 2011 Indianapolis Belmont WWTP, Citizens Thermal, and IP&L Harding Street point source emissions, locations, and stack parameters.

Facility Name	AERMOD source ID	Emissions (tons $year^{-1}$)	Emission factor	UTM-x (m)	UTM-y (m)	Elevation (m)	Stack height (m)	Stack temperature (K)	Stack velocity (m s^{-1})	Stack diameter (m)
INDIANAPOLIS BELMONT WWTP	BELL_SN1	24.90	MONTH	568970.00	4397879.00	208.61	45.72	297.59	0.64	3.20
Citizens Thermal	CIT_SN1	2254.90	MONTH	571351.00	4401766.00	216.45	82.91	566.48	4.60	4.42
Citizens Thermal	CIT_SN2	2093.70	MONTH	571396.00	4401766.00	217.46	82.91	463.71	4.72	4.64
Citizens Thermal	CIT_SN3	0.16	MONTH	571380.00	4401766.00	217.35	82.91	488.71	5.33	4.63
Citizens Thermal	CIT_SN4	0.08	MONTH	571396.00	4401766.00	217.46	82.91	463.71	4.72	4.64
Citizens Thermal	CIT_SN5	0.0005	MONTH	571380.00	4401766.00	217.35	82.91	488.71	5.33	4.63
IP&L - HARDING STREET	IPL_SE1	0.11	HOURLY	569200.00	4396339.00	208.02	9.45	791.48	7.16	3.81
IP&L - HARDING STREET	IPL_SE2	0.10	HOURLY	569180.00	4396327.00	207.98	9.75	791.48	7.16	3.81
IP&L - HARDING STREET	IPL_SE3	0.10	HOURLY	568867.00	4396303.00	208.00	20.12	827.59	57.39	4.21
IP&L - HARDING STREET	IPL_SE4	8633.50	HOURLY	568749.00	4396008.00	208.08	79.55	440.93	65.84	1.98
IP&L - HARDING STREET	IPL_SE5	7940.50	HOURLY	568752.00	4395965.00	208.32	79.55	449.82	63.52	1.98
IP&L - HARDING STREET	IPL_SE6	680.70	HOURLY	568984.00	4395792.00	206.56	172.21	329.26	14.33	6.10
IP&L - HARDING STREET	IPL_SE7	1739.00	HOURLY	568984.00	4395792.00	206.56	172.21	414.82	23.44	6.10
IP&L - HARDING STREET	IPL_SE8	0.20	HOURLY	569050.00	4396339.00	208.26	22.86	810.93	36.58	5.49

Table B-18. 2011 Rolls Royce, Vertellus, and Quemetco point source emissions, locations, and stack parameters.

Facility Name	AERMOD source ID	Emissions (tons year^{-1})	Emission factor	UTM-x (m)	UTM-y (m)	Elevation (m)	Stack height (m)	Stack temperature (K)	Stack velocity (m s^{-1})	Stack diameter (m)
ROLLS ROYCE	RR_SN1	0.02	MONTH	567493.00	4398570.00	212.29	4.57	866.48	32.34	0.30
ROLLS ROYCE	RR_SN2	0.89	HROFDY	567428.00	4398870.00	212.70	17.37	588.71	33.04	1.22
ROLLS ROYCE	RR_SN3	16.17	HROFDY	567402.00	4398886.00	212.70	19.81	755.37	45.51	0.91
ROLLS ROYCE	RR_SN9	3.60	MHRDOW	567435.00	4398899.00	212.72	9.14	866.48	21.21	1.52
ROLLS ROYCE	RR_SN10	6.19	MONTH	567551.00	4399165.00	212.10	15.24	677.59	17.47	1.98
ROLLS ROYCE	RR_SN11	0.06	MONTH	567551.00	4399165.00	212.10	15.24	677.59	17.47	1.98
ROLLS ROYCE	RR_SN12	23.36	MONTH	567544.50	4399165.00	212.24	15.24	677.59	17.47	1.98
ROLLS ROYCE	RR_SN13	1.56	MONTH	567512.00	4399163.00	212.51	18.29	533.15	6.52	1.22
ROLLS ROYCE	RR_SN14	0.04	MHRDOW	567513.00	4399174.00	212.61	9.14	866.48	21.21	1.52
ROLLS ROYCE	RR_SN15	0.002	MONTH	567439.00	4398911.00	212.70	15.24	755.37	13.53	1.68
VERTELLUS	VERT_SN1	3.98	MONTH	566836.00	4399683.00	214.94	9.14	453.71	6.28	1.22
VERTELLUS	VERT_SN2	26.43	MONTH	566981.00	4399746.00	215.16	9.14	504.26	7.53	1.22
VERTELLUS	VERT_SN4	0.19	MONTH	566995.00	4399731.00	214.89	10.97	422.04	5.49	0.81
VERTELLUS	VERT_SN12	0.04	MONTH	566851.06	4399666.50	214.85	20.42	823.15	5.09	1.07
VERTELLUS	VERT_SN13	0.02	MONTH	566901.00	4399710.00	215.15	20.73	823.15	5.09	1.07
VERTELLUS	VERT_SN14	0.05	MONTH	566866.94	4399637.00	214.79	21.64	633.15	6.10	1.52
VERTELLUS	VERT_SN15	0.03	MONTH	566864.94	4399640.00	214.79	24.69	823.15	6.07	1.07
QUEMETCO	QUE_SN1	70.78	MONTH	559977.54	4400993.45	235.78	30.48	327.04	16.86	1.22
QUEMETCO	QUE_SN2	53.59	MONTH	559993.31	4400853.53	235.10	50.29	321.48	14.84	3.35

Table B-19. 2011 Indianapolis, IN Rolls Royce area source emissions, locations, and release parameters.

Facility Name	AERMOD source ID	Emissions (tons year^{-1})	Emission factor	UTM-x (m)	UTM-y (m)	Elevation (m)	Release height (m)	X-dimension (m)	Y-dimension (m)	Angle	σ_z (m)
ROLLS ROYCE	RR_SN4	1.30	HROFDY	567593.31	4398478.50	211.876	3.05	10	10	0	0
ROLLS ROYCE	RR_SN5	0.33	HROFDY	567359.62	4398742.50	212.987	3.05	10	10	0	0
ROLLS ROYCE	RR_SN6	4.58	HROFDY	567492.69	4399179.00	212.8	3.05	10	10	0	0
ROLLS ROYCE	RR_SN7	0.0003	MONTH	567593.31	4398478.50	211.876	3.05	10	10	0	0
ROLLS ROYCE	RR_SN8	0.0007	MONTH	567492.69	4399179.00	212.8	3.05	10	10	0	0

Table B-20. 2012 Indianapolis Belmont WWTP, Citizens Thermal, and IP&L Harding Street point source emissions, locations, and stack parameters.

Facility Name	AERMOD source ID	Emissions (tons year^{-1})	Emission factor	UTM-x (m)	UTM-y (m)	Elevation (m)	Stack height (m)	Stack temperature (K)	Stack velocity (m s^{-1})	Stack diameter (m)
INDIANAPOLIS BELMONT WWTP	BELL_SN1	24.90	MONTH	568970.00	4397879.00	208.61	45.72	297.59	0.64	3.20
Citizens Thermal	CIT_SN1	2002.70	MONTH	571351.00	4401766.00	216.45	82.91	566.48	4.60	4.42
Citizens Thermal	CIT_SN2	1849.50	MONTH	571396.00	4401766.00	217.46	82.91	463.71	4.72	4.64
Citizens Thermal	CIT_SN3	0.0000004	MONTH	571351.00	4401766.00	216.45	82.91	566.48	4.60	4.42
Citizens Thermal	CIT_SN4	0.18	MONTH	571380.00	4401766.00	217.35	82.91	488.71	5.33	4.63
Citizens Thermal	CIT_SN5	0.07	MONTH	571396.00	4401766.00	217.46	82.91	463.71	4.72	4.64
Citizens Thermal	CIT_SN6	0.001	MONTH	571380.00	4401766.00	217.35	82.91	488.71	5.33	4.63
IP&L - HARDING STREET	IPL_SE1	0.19	HOURLY	569200.00	4396339.00	208.02	9.45	791.48	7.16	3.81
IP&L - HARDING STREET	IPL_SE2	0.16	HOURLY	569180.00	4396327.00	207.98	9.75	791.48	7.16	3.81
IP&L - HARDING STREET	IPL_SE3	10531.00	HOURLY	568749.00	4396008.00	208.08	79.55	440.93	65.84	1.98
IP&L - HARDING STREET	IPL_SE4	10270.00	HOURLY	568752.00	4395965.00	208.32	79.55	449.82	63.52	1.98
IP&L - HARDING STREET	IPL_SE5	632.10	HOURLY	568984.00	4395792.00	206.56	172.21	329.26	14.33	6.10
IP&L - HARDING STREET	IPL_SE6	109.00	HOURLY	568984.00	4395792.00	206.56	172.21	414.82	23.44	6.10
IP&L - HARDING STREET	IPL_SE7	0.20	HOURLY	569050.00	4396339.00	208.26	22.86	810.93	36.58	5.49

Table B-21. 2012 Rolls Royce, Vertellus, and Quemetco point source emissions, locations, and stack parameters.

Facility Name	AERMOD source ID	Emissions (tons year^{-1})	Emission factor	UTM-x (m)	UTM-y (m)	Elevation (m)	Stack height (m)	Stack temperature (K)	Stack velocity (m s^{-1})	Stack diameter (m)
ROLLS ROYCE	RR_SN1	1.74	MONTH	567493.00	4398570.00	212.29	4.57	866.48	32.34	0.30
ROLLS ROYCE	RR_SN2	0.23	HROFDY	567402.00	4398886.00	212.70	19.81	755.37	45.51	0.91
ROLLS ROYCE	RR_SN3	0.70	HROFDY	567428.00	4398870.00	212.70	17.37	588.71	33.04	1.22
ROLLS ROYCE	RR_SN4	13.98	HROFDY	567402.00	4398886.00	212.70	19.81	755.37	45.51	0.91
ROLLS ROYCE	RR_SN9	3.49	MHRDOW	567435.00	4398899.00	212.72	9.14	866.48	21.21	1.52
ROLLS ROYCE	RR_SN10	6.08	MONTH	567551.00	4399165.00	212.10	15.24	677.59	17.47	1.98
ROLLS ROYCE	RR_SN11	0.03	MONTH	567551.00	4399165.00	212.10	15.24	677.59	17.47	1.98
ROLLS ROYCE	RR_SN12	7.29	MONTH	567544.50	4399165.00	212.24	15.24	677.59	17.47	1.98
ROLLS ROYCE	RR_SN13	1.57	MONTH	567512.00	4399163.00	212.51	18.29	533.15	6.52	1.22
ROLLS ROYCE	RR_SN14	0.02	MHRDOW	567513.00	4399174.00	212.61	9.14	866.48	21.21	1.52
ROLLS ROYCE	RR_SN15	0.0001	MONTH	567439.00	4398911.00	212.70	15.24	755.37	13.53	1.68
VERTELLUS	VERT_SN1	1.38	MONTH	566836.00	4399683.00	214.94	9.14	453.71	6.28	1.22
VERTELLUS	VERT_SN2	22.18	MONTH	566981.00	4399746.00	215.16	9.14	504.26	7.53	1.22
VERTELLUS	VERT_SN4	0.90	MONTH	566995.00	4399731.00	214.89	10.97	422.04	5.49	0.81
VERTELLUS	VERT_SN12	0.06	MONTH	566851.06	4399666.50	214.85	20.42	823.15	5.09	1.07
VERTELLUS	VERT_SN13	0.01	MONTH	566901.00	4399710.00	215.15	20.73	823.15	5.09	1.07
VERTELLUS	VERT_SN14	0.03	MONTH	566866.94	4399637.00	214.79	21.64	633.15	6.10	1.52
VERTELLUS	VERT_SN15	0.02	MONTH	566864.94	4399640.00	214.79	24.69	823.15	6.07	1.07
QUEMETCO	QUE_SN1	70.78	MONTH	559977.54	4400993.45	235.78	30.48	327.04	16.86	1.22
QUEMETCO	QUE_SN2	53.59	MONTH	559993.31	4400853.53	235.10	50.29	321.48	14.84	3.35

Table B-22. 2012 Indianapolis, IN Rolls Royce area source emissions, locations, and release parameters.

Facility Name	AERMOD source ID	Emissions (tons year-1)	Emission factor	UTM-x (m)	UTM-y (m)	Elevation (m)	Release height (m)	X-dimension (m)	Y-dimension (m)	Angle	σ_z (m)
ROLLS ROYCE	RR_SN5	1.33	HROFDY	567593.31	4398478.50	211.88	3.05	10	10	0	0
ROLLS ROYCE	RR_SN6	0.47	HROFDY	567359.63	4398742.50	212.99	3.05	10	10	0	0
ROLLS ROYCE	RR_SN7	2.49	HROFDY	567492.69	4399179.00	212.80	3.05	10	10	0	0
ROLLS ROYCE	RR_SN8	0.001	MONTH	567492.69	4399179.00	212.80	3.05	10	10	0	0

Table B-23. 2013 Indianapolis Belmont WWTP, Citizens Thermal, and IP&L Harding Street point source emissions, locations, and stack parameters.

Facility Name	AERMOD source ID	Emissions (tons $year^{-1}$)	Emission factor	UTM-x (m)	UTM-y (m)	Elevation (m)	Stack height (m)	Stack temperature (K)	Stack velocity (m s^{-1})	Stack diameter (m)
INDIANAPOLIS BELMONT WWTP	BELL_SN1	20.10	MONTH	568970.00	4397879.00	208.61	45.72	297.59	0.64	3.20
Citizens Thermal	CIT_SN1	2229.80	MONTH	571351.00	4401766.00	216.45	82.91	566.48	4.60	4.42
Citizens Thermal	CIT_SN2	1575.00	MONTH	571396.00	4401766.00	217.46	82.91	463.71	4.72	4.64
Citizens Thermal	CIT_SN3	0.0000001	MONTH	571351.00	4401766.00	216.45	82.91	566.48	4.60	4.42
Citizens Thermal	CIT_SN4	0.28	MONTH	571380.00	4401766.00	217.35	82.91	488.71	5.33	4.63
Citizens Thermal	CIT_SN5	0.24	MONTH	571396.00	4401766.00	217.46	82.91	463.71	4.72	4.64
Citizens Thermal	CIT_SN6	0.002	MONTH	571380.00	4401766.00	217.35	82.91	488.71	5.33	4.63
IP&L - HARDING STREET	IPL_SE1	0.02	HOURLY	569200.00	4396339.00	208.02	9.45	791.48	7.16	3.81
IP&L - HARDING STREET	IPL_SE2	0.01	HOURLY	569180.00	4396327.00	207.98	9.75	791.48	7.16	3.81
IP&L - HARDING STREET	IPL_SE3	0.20	HOURLY	568867.00	4396303.00	208.00	20.12	827.59	57.39	4.21
IP&L - HARDING STREET	IPL_SE4	0.20	HOURLY	568910.00	4396306.00	208.01	20.12	822.04	62.15	4.21
IP&L - HARDING STREET	IPL_SE5	13324.00	HOURLY	568749.00	4396008.00	208.08	79.55	440.93	65.84	1.98
IP&L - HARDING STREET	IPL_SE6	12603.00	HOURLY	568752.00	4395965.00	208.32	79.55	449.82	63.52	1.98
IP&L - HARDING STREET	IPL_SE7	1846.10	HOURLY	568984.00	4395792.00	206.56	172.21	329.26	14.33	6.10
IP&L - HARDING STREET	IPL_SE8	200.30	HOURLY	568984.00	4395792.00	206.56	172.21	414.82	23.44	6.10
IP&L - HARDING STREET	IPL_SE9	0.30	HOURLY	569050.00	4396339.00	208.26	22.86	810.93	36.58	5.49

Table B-24. 2013 Rolls Royce, Vertellus, and Quemetco point source emissions, locations, and stack parameters.

Facility Name	AERMOD source ID	Emissions (tons year^{-1})	Emission factor	UTM-x (m)	UTM-y (m)	Elevation (m)	Stack height (m)	Stack temperature (K)	Stack velocity (m s^{-1})	Stack diameter (m)
ROLLS ROYCE	RR_SN1	0.48	MONTH	567493.00	4398570.00	212.29	4.57	866.48	32.34	0.30
ROLLS ROYCE	RR_SN3	1.15	HROFDY	567428.00	4398870.00	212.70	17.37	588.71	33.04	1.22
ROLLS ROYCE	RR_SN4	12.39	HROFDY	567402.00	4398886.00	212.70	19.81	755.37	45.51	0.91
ROLLS ROYCE	RR_SN9	2.78	MHRDOW	567435.00	4398899.00	212.72	9.14	866.48	21.21	1.52
ROLLS ROYCE	RR_SN10	7.24	MONTH	567551.00	4399165.00	212.10	15.24	677.59	17.47	1.98
ROLLS ROYCE	RR_SN11	0.05	MONTH	567551.00	4399165.00	212.10	15.24	677.59	17.47	1.98
ROLLS ROYCE	RR_SN12	2.80	MONTH	567544.50	4399165.00	212.24	15.24	677.59	17.47	1.98
ROLLS ROYCE	RR_SN13	4.77	MONTH	567512.00	4399163.00	212.51	18.29	533.15	6.52	1.22
ROLLS ROYCE	RR_SN14	0.02	MHRDOW	567513.00	4399174.00	212.61	9.14	866.48	21.21	1.52
ROLLS ROYCE	RR_SN15	0.001	MONTH	567439.00	4398911.00	212.70	15.24	755.37	13.53	1.68
VERTELLUS	VERT_SN1	25.01	MONTH	566981.00	4399746.00	215.16	9.14	504.26	7.53	1.22
VERTELLUS	VERT_SN10	0.07	MONTH	566851.06	4399666.50	214.85	20.42	823.15	5.09	1.07
VERTELLUS	VERT_SN11	0.02	MONTH	566901.00	4399710.00	215.15	20.73	823.15	5.09	1.07
VERTELLUS	VERT_SN12	0.02	MONTH	566866.94	4399637.00	214.79	21.64	633.15	6.10	1.52
VERTELLUS	VERT_SN13	0.02	MONTH	566864.94	4399640.00	214.79	24.69	823.15	6.07	1.07
QUEMETCO	QUE_SN1	68.77	MONTH	559977.54	4400993.45	235.78	30.48	327.04	16.86	1.22
QUEMETCO	QUE_SN2	3.97	MONTH	559993.31	4400853.53	235.10	50.29	297.04	10.70	3.35
QUEMETCO	QUE_SN3	23.82	MONTH	559993.31	4400853.53	235.10	50.29	321.48	14.84	3.35

Table B-25. 2013 Rolls Royce area source emissions, locations, and release parameters.

Facility Name	AERMOD source ID	Emissions (tons year^{-1})	Emission factor	UTM-x (m)	UTM-y (m)	Elevation (m)	Release height (m)	X-dimension (m)	Y-dimension (m)	Angle	σ_z (m)
ROLLS ROYCE	RR_SN2	0.001	HROFDY	567492.69	4399179.00	212.80	3.05	10	10	0	0
ROLLS ROYCE	RR_SN5	1.39	HROFDY	567593.31	4398478.50	211.88	3.05	10	10	0	0
ROLLS ROYCE	RR_SN6	0.61	HROFDY	567359.63	4398742.50	212.99	3.05	10	10	0	0
ROLLS ROYCE	RR_SN7	3.02	HROFDY	567492.69	4399179.00	212.80	3.05	10	10	0	0

Table B-26. Tulsa Refinery-East 2011-2013 AERMOD source identifier crosswalk.

Facility Name	Unit ID	Process ID	Release Point ID	AERMOD 2011	AERMOD 2012	AERMOD 2013
TULSA RFNRY-EAST	5070713	4882912	15790214	REFEAST_SN1	REFEAST_SN1	REFEAST_SN1
TULSA RFNRY-EAST	72309613	66435812	100082714	REFEAST_SN2	REFEAST_SN2	REFEAST_SN2
TULSA RFNRY-EAST	5070913	4882512	15790014	REFEAST_SN4	REFEAST_SN4	REFEAST_SN4
TULSA RFNRY-EAST	5070813	4882812	15790114	REFEAST_SN5	REFEAST_SN5	REFEAST_SN5
TULSA RFNRY-EAST	72309513	66435712	100082614	REFEAST_SN6	REFEAST_SN6	REFEAST_SN6
TULSA RFNRY-EAST	72309413	66435612	100082514	REFEAST_SN7	REFEAST_SN7	REFEAST_SN7
TULSA RFNRY-EAST	72308613	66437212	100081514	REFEAST_SN8	REFEAST_SN8	REFEAST_SN8
TULSA RFNRY-EAST	5070213	4883812	15790914	REFEAST_SN9	REFEAST_SN9	REFEAST_SN9
TULSA RFNRY-EAST	5066913	4883712	15659614	REFEAST_SN11	REFEAST_SN11	REFEAST_SN11
TULSA RFNRY-EAST	5070613	4882412	15790314	REFEAST_SN12	REFEAST_SN12	REFEAST_SN12
TULSA RFNRY-EAST	72310013	66436712	100083414	REFEAST_SN14	REFEAST_SN14	REFEAST_SN14
TULSA RFNRY-EAST	5064513	4880712	15786414	REFEAST_SN15	REFEAST_SN15	REFEAST_SN15
TULSA RFNRY-EAST	5064313	4884112	15786714	REFEAST_SN16	REFEAST_SN16	REFEAST_SN16
TULSA RFNRY-EAST	5071113	4882612	15789714	REFEAST_SN17	REFEAST_SN18	REFEAST_SN17
TULSA RFNRY-EAST	5071913	4883912	15788214	REFEAST_SN18	REFEAST_SN19	REFEAST_SN18

Table B-27. Tulsa Refinery-West 2011-2013 AERMOD source identifier crosswalk.

Facility Name	Unit ID	Process ID	Release Point ID	AERMOD 2011	AERMOD 2012	AERMOD 2013
TULSA RFNRY WEST	72317213	66440812	100094814	REFWEST_SN1	REFWEST_SN1	
TULSA RFNRY WEST	651913	655212	15606514	REFWEST_SN2	REFWEST_SN2	
TULSA RFNRY WEST	652113	657112	15606314	REFWEST_SN3	REFWEST_SN3	REFWEST_SN1
TULSA RFNRY WEST	663913	654212	16298114			REFWEST_SN2
TULSA RFNRY WEST	72311713	66439312	100085314	REFWEST_SN4	REFWEST_SN4	
TULSA RFNRY WEST	664013	654712	16298014			REFWEST_SN3
TULSA RFNRY WEST	660813	654812	16303714			REFWEST_SN4
TULSA RFNRY WEST	107042213	110579312	151543514		REFWEST_SN5	REFWEST_SN5
TULSA RFNRY WEST	654413	660012	15477714	REFWEST_SN5	REFWEST_SN6	REFWEST_SN6
TULSA RFNRY WEST	654313	659912	15477814	REFWEST_SN6	REFWEST_SN7	REFWEST_SN7
TULSA RFNRY WEST	654613	655312	15477514	REFWEST_SN7	REFWEST_SN8	REFWEST_SN8
TULSA RFNRY WEST	663113	651512	16299414	REFWEST_SN8	REFWEST_SN9	REFWEST_SN9
TULSA RFNRY WEST	651113	662812	15607614	REFWEST_SN9	REFWEST_SN11	REFWEST_SN11
TULSA RFNRY WEST	650813	656012	15607914	REFWEST_SN10	REFWEST_SN10	REFWEST_SN10
TULSA RFNRY WEST	653513	659612	15478614	REFWEST_SN11	REFWEST_SN12	REFWEST_SN12
TULSA RFNRY WEST	654213	662012	15477914	REFWEST_SN12	REFWEST_SN13	REFWEST_SN13
TULSA RFNRY WEST	651013	658912	15607714	REFWEST_SN13	REFWEST_SN14	REFWEST_SN14
TULSA RFNRY WEST	653613	659012	15478514	REFWEST_SN14	REFWEST_SN15	REFWEST_SN15
TULSA RFNRY WEST	651713	655012	15606714	REFWEST_SN15	REFWEST_SN16	REFWEST_SN16
TULSA RFNRY WEST	654113	663512	15478014	REFWEST_SN16	REFWEST_SN17	REFWEST_SN17
TULSA RFNRY WEST	651313	658812	15607314	REFWEST_SN17	REFWEST_SN18	REFWEST_SN18
TULSA RFNRY WEST	650913	654912	15607814	REFWEST_SN18	REFWEST_SN19	REFWEST_SN19
TULSA RFNRY WEST	651413	661212	15607214	REFWEST_SN19	REFWEST_SN20	REFWEST_SN20
TULSA RFNRY WEST	663813	651712	16298214			REFWEST_SN21
TULSA RFNRY WEST	654513	656112	15477614			REFWEST_SN22
TULSA RFNRY WEST	653713	659112	15478414			REFWEST_SN23
TULSA RFNRY WEST	658713	651412	16408914	REFWEST_SN20	REFWEST_SN21	REFWEST_SN24

Table B-28. PSO Northeastern Power Station and Sapulpa 2011-2013 AERMOD source identifier crosswalk.

Facility Name	Unit ID	Process ID	Release Point ID	AERMOD 2011	AERMOD 2012	AERMOD 2013
PSO NORTHEASTERN	6698313	6664412	15999814	PSO_SE1	PSO_SE1	PSO_SE1
PSO NORTHEASTERN	6698313	6664412	15999914	PSO_SE2	PSO_SE2	PSO_SE2
PSO NORTHEASTERN	6698513	6664212	15999514	PSO_SE3	PSO_SE3	PSO_SE3
PSO NORTHEASTERN	6698813	6664412	15999114	PSO_SE4	PSO_SE4	PSO_SE4
PSO NORTHEASTERN	6698813	6664412	15999214	PSO_SE5	PSO_SE5	PSO_SE5
PSO NORTHEASTERN	6698913	6664012	15999014	PSO_SE6	PSO_SE6	PSO_SE6
PSO NORTHEASTERN	6699113	6664712	15998814	PSO_SE7	PSO_SE7	PSO_SE7
SAPULPA	8331213	8217212	17068814	SAP_SN1	SAP_SN1	SAP_SN2
SAPULPA	8331413	8217312	17068514	SAP_SN2	SAP_SN2	SAP_SN3
SAPULPA	72251213	66374812	100009614	SAP_SN3	SAP_SN3	SAP_SN4
SAPULPA	8331113	66375112	17068914	SAP_SN4	SAP_SN4	SAP_SN5
SAPULPA	8331113	66375212	17068914	SAP_SN4	SAP_SN4	SAP_SN5
SAPULPA	8331113	66375312	17068914	SAP_SN4	SAP_SN4	SAP_SN5
SAPULPA	8331113	66375412	17068914	SAP_SN4	SAP_SN4	SAP_SN5
SAPULPA	8331113	66375512	17068914	SAP_SN4	SAP_SN4	SAP_SN5
SAPULPA	8331113	66375612	17068914	SAP_SN4	SAP_SN4	SAP_SN5
SAPULPA	8331113	66375712	17068914	SAP_SN4	SAP_SN4	SAP_SN5
SAPULPA	8331113	66376612	17068914	SAP_SN4	SAP_SN4	SAP_SN5
SAPULPA	8331113	66375812	17068914	SAP_SN5	SAP_SN5	SAP_SN6
SAPULPA	8331113	66375912	17068914	SAP_SN5	SAP_SN5	SAP_SN6
SAPULPA	8331113	66376012	17068914	SAP_SN5	SAP_SN5	SAP_SN6
SAPULPA	8331113	66376112	17068914	SAP_SN5	SAP_SN5	SAP_SN6
SAPULPA	8331113	66376212	17068914	SAP_SN5	SAP_SN5	SAP_SN6
SAPULPA	8331113	66376312	17068914	SAP_SN5	SAP_SN5	SAP_SN6
SAPULPA	8331113	66376412	17068914	SAP_SN5	SAP_SN5	SAP_SN6
SAPULPA	8331113	66376512	17068914	SAP_SN5	SAP_SN5	SAP_SN6
SAPULPA	72251313	66375012	100009714	SAP_SN6	SAP_SN6	SAP_SN7
SAPULPA	108757113	112230012	153985314			SAP_SN1

Table B-29. 2011 Tulsa East Refinery point source emissions, locations, and stack parameters.

Facility Name	AERMOD source ID	Emissions (tons year^{-1})	Emission factor	UTM-x (m)	UTM-y (m)	Elevation (m)	Stack height (m)	Stack temperature (K)	Stack velocity (m s^{-1})	Stack diameter (m)
TULSA RFNRY-EAST	REFEAST_SN1	2.00	MONTH	230409.02	4000701.87	192.12	73.15	1088.71	43.34	0.49
TULSA RFNRY-EAST	REFEAST_SN2	0.25	MONTH	229761.77	4000607.68	192.00	30.78	317.59	6.68	0.76
TULSA RFNRY-EAST	REFEAST_SN4	0.83	MONTH	229823.09	4000610.90	192.00	60.96	444.26	5.88	0.61
TULSA RFNRY-EAST	REFEAST_SN5	15.21	MONTH	229944.74	4000860.87	194.00	58.22	572.59	22.92	1.52
TULSA RFNRY-EAST	REFEAST_SN6	0.12	MONTH	229658.38	4000653.14	192.00	29.26	313.71	8.23	1.13
TULSA RFNRY-EAST	REFEAST_SN7	0.13	MONTH	229663.74	4000658.82	192.00	30.48	311.48	7.86	1.13
TULSA RFNRY-EAST	REFEAST_SN8	0.04	MONTH	229954.38	4001000.54	192.90	13.72	570.37	16.06	1.07
TULSA RFNRY-EAST	REFEAST_SN9	0.25	MONTH	229946.71	4000617.28	194.83	42.67	583.15	14.60	1.46
TULSA RFNRY-EAST	REFEAST_SN11	0.44	MONTH	229945.36	4000870.85	194.00	46.02	624.82	3.99	1.77
TULSA RFNRY-EAST	REFEAST_SN12	1.83	MONTH	229956.32	4001096.60	194.47	53.34	449.82	3.47	3.51
TULSA RFNRY-EAST	REFEAST_SN14	0.66	MONTH	229971.12	4000687.91	194.77	38.10	466.48	3.84	2.53
TULSA RFNRY-EAST	REFEAST_SN15	0.74	MONTH	229950.17	4000673.17	194.97	37.80	560.93	10.00	1.77
TULSA RFNRY-EAST	REFEAST_SN16	0.16	MONTH	229950.84	4000700.18	195.00	37.80	533.15	7.04	1.37
TULSA RFNRY-EAST	REFEAST_SN17	1.44	MONTH	229912.85	4001441.17	192.39	21.64	449.82	6.25	2.13
TULSA RFNRY-EAST	REFEAST_SN18	1.43	MONTH	229940.84	4001441.17	192.31	21.64	449.82	6.19	2.13

Table B-30. 2011 Tulsa West Refinery point source emissions, locations, and stack parameters.

Facility Name	AERMOD source ID	Emissions (tons year^{-1})	Emission factor	UTM-x (m)	UTM-y (m)	Elevation (m)	Stack height (m)	Stack temperature (K)	Stack velocity (m s^{-1})	Stack diameter (m)
TULSA RFNRY WEST	REFWEST_SN1	0.03	MONTH	228617.00	4003889.00	195.00	5.49	616.48	5.06	0.15
TULSA RFNRY WEST	REFWEST_SN2	0.005	MONTH	228750.30	4003806.26	195.10	6.71	588.71	13.20	0.15
TULSA RFNRY WEST	REFWEST_SN3	5.73	MONTH	228706.00	4002861.00	195.00	43.89	477.59	11.83	0.30
TULSA RFNRY WEST	REFWEST_SN4	0.007	MONTH	228658.38	4003859.03	195.10	7.62	547.04	7.25	0.21
TULSA RFNRY WEST	REFWEST_SN5	44.78	MONTH	229176.29	4003711.77	195.10	30.48	637.59	1.92	1.62
TULSA RFNRY WEST	REFWEST_SN6	380.27	MONTH	229185.32	4003728.24	195.10	38.10	548.15	5.15	1.62
TULSA RFNRY WEST	REFWEST_SN7	103.02	MONTH	229202.04	4003723.20	195.20	18.90	505.93	2.99	1.07
TULSA RFNRY WEST	REFWEST_SN8	866.22	MONTH	228262.29	4003837.45	194.30	41.15	522.04	4.88	2.26
TULSA RFNRY WEST	REFWEST_SN9	39.26	MONTH	228236.99	4003995.32	194.20	15.24	471.48	4.11	0.85
TULSA RFNRY WEST	REFWEST_SN10	37.86	MONTH	228237.62	4003989.27	194.20	15.24	683.15	2.99	1.37
TULSA RFNRY WEST	REFWEST_SN11	0.006	MONTH	228251.07	4004028.52	193.90	25.91	768.71	4.05	1.52
TULSA RFNRY WEST	REFWEST_SN12	0.01	MONTH	228262.17	4004029.83	193.90	27.43	736.48	2.19	2.13
TULSA RFNRY WEST	REFWEST_SN13	157.00	MONTH	228246.58	4004020.78	193.90	27.74	922.04	4.82	2.13
TULSA RFNRY WEST	REFWEST_SN14	18.64	MONTH	228246.08	4004012.79	193.90	30.78	877.59	2.04	1.13
TULSA RFNRY WEST	REFWEST_SN15	59.37	MONTH	228239.18	4003982.16	194.30	23.47	523.15	26.33	0.61
TULSA RFNRY WEST	REFWEST_SN16	36.35	MONTH	229175.91	4003721.81	195.10	27.43	560.93	3.20	0.91
TULSA RFNRY WEST	REFWEST_SN17	43.23	MONTH	228239.37	4003969.12	194.60	20.12	594.26	2.38	1.37
TULSA RFNRY WEST	REFWEST_SN18	74.03	MONTH	228279.45	4003823.37	194.50	38.10	726.48	2.26	2.13
TULSA RFNRY WEST	REFWEST_SN19	270.43	MONTH	228279.45	4003823.37	194.50	38.10	738.71	4.88	2.26
TULSA RFNRY WEST	REFWEST_SN20	460.16	MONTH	228688.88	4003894.68	195.19	33.53	394.26	3.41	3.20

Table B-31. 2011 PSO Northeastern and Sapulpa point source emissions, locations, and stack parameters.

Facility Name	AERMOD source ID	Emissions (tons year^{-1})	Emission factor	UTM-x (m)	UTM-y (m)	Elevation (m)	Stack height (m)	Stack temperature (K)	Stack velocity (m s^{-1})	Stack diameter (m)
PSO NORTHEASTERN	PSO_SE1	9007.70	HOURLY	258002.59	4034618.88	195.67	182.88	419.26	13.81	8.23
PSO NORTHEASTERN	PSO_SE2	26.14	HOURLY	258002.59	4034618.88	195.67	182.88	419.26	13.81	8.23
PSO NORTHEASTERN	PSO_SE3	2.36	HOURLY	257841.41	4035283.44	195.41	55.78	393.71	16.28	5.49
PSO NORTHEASTERN	PSO_SE4	8879.30	HOURLY	258002.59	4034618.88	195.67	182.88	419.26	13.81	8.23
PSO NORTHEASTERN	PSO_SE5	25.54	HOURLY	258002.59	4034618.88	195.67	182.88	419.26	13.81	8.23
PSO NORTHEASTERN	PSO_SE6	0.18	HOURLY	257850.92	4035160.78	195.23	45.72	366.48	19.69	5.74
PSO NORTHEASTERN	PSO_SE7	0.20	HOURLY	257850.92	4035160.78	195.23	45.72	366.48	21.55	5.49
SAPULPA	SAP_SN1	100.32	MONTH	220648.04	3989373.19	215.01	28.35	530.37	9.60	1.86
SAPULPA	SAP_SN2	33.08	MONTH	220621.83	3989378.25	215.62	32.31	498.71	19.39	1.29
SAPULPA	SAP_SN3	78.85	MONTH	220621.83	3989378.25	215.62	29.87	515.37	10.27	1.71
SAPULPA	SAP_SN4	0.02	MONTH	220667.19	3989381.92	214.54	26.52	310.93	2.13	2.29
SAPULPA	SAP_SN5	0.03	MONTH	220667.19	3989381.92	214.54	29.26	310.93	2.13	2.29

Table B-32. 2011 Sapulpa area source emissions, locations, and release parameters.

Facility Name	AERMOD source ID	Emissions (tons year^{-1})	Emission factor	UTM-x (m)	UTM-y (m)	Elevation (m)	Release height (m)	X-dimension (m)	Y-dimension (m)	Angle	σ_z (m)
SAPULPA	SAP_SN6	0.03	MONTH	220691.84	3989080	218.06	10.67	2.74	2.74	0	2.48

Table B-33. 2012 Tulsa East Refinery point source emissions, locations, and stack parameters.

Facility Name	AERMOD source ID	Emissions (tons year^{-1})	Emission factor	UTM-x (m)	UTM-y (m)	Elevation (m)	Stack height (m)	Stack temperature (K)	Stack velocity (m s^{-1})	Stack diameter (m)
TULSA RFNRY-EAST	REFEAST_SN1	3.91	MONTH	230409.02	4000701.87	192.12	73.15	1088.71	43.34	0.49
TULSA RFNRY-EAST	REFEAST_SN2	1.15	MONTH	229761.77	4000607.68	192.00	30.78	317.59	6.68	0.76
TULSA RFNRY-EAST	REFEAST_SN4	0.38	MONTH	229823.09	4000610.90	192.00	60.96	444.26	5.88	0.61
TULSA RFNRY-EAST	REFEAST_SN5	11.19	MONTH	229944.74	4000860.87	194.00	58.22	572.59	22.92	1.52
TULSA RFNRY-EAST	REFEAST_SN6	0.14	MONTH	229658.38	4000653.14	192.00	29.26	313.71	8.23	1.13
TULSA RFNRY-EAST	REFEAST_SN7	0.15	MONTH	229663.74	4000658.82	192.00	30.48	311.48	7.86	1.13
TULSA RFNRY-EAST	REFEAST_SN8	0.04	MONTH	229954.38	4001000.54	192.90	13.72	570.37	16.06	1.07
TULSA RFNRY-EAST	REFEAST_SN9	0.26	MONTH	229946.71	4000617.28	194.83	42.67	583.15	14.60	1.46
TULSA RFNRY-EAST	REFEAST_SN11	0.58	MONTH	229945.36	4000870.85	194.00	46.02	624.82	3.99	1.77
TULSA RFNRY-EAST	REFEAST_SN12	1.35	MONTH	229956.32	4001096.60	194.47	53.34	449.82	3.47	3.51
TULSA RFNRY-EAST	REFEAST_SN14	0.62	MONTH	229971.12	4000687.91	194.77	38.10	466.48	3.84	2.53
TULSA RFNRY-EAST	REFEAST_SN15	0.72	MONTH	229950.17	4000673.17	194.97	37.80	560.93	10.00	1.77
TULSA RFNRY-EAST	REFEAST_SN16	0.16	MONTH	229950.84	4000700.18	195.00	37.80	533.15	7.04	1.37
TULSA RFNRY-EAST	REFEAST_SN18	1.46	MONTH	229912.85	4001441.17	192.39	21.64	449.82	6.25	2.13
TULSA RFNRY-EAST	REFEAST_SN19	1.20	MONTH	229940.84	4001441.17	192.31	21.64	449.82	6.19	2.13

Table B-34. 2012 Tulsa West Refinery point source emissions, locations, and stack parameters.

Facility Name	AERMOD source ID	Emissions (tons year^{-1})	Emission factor	UTM-x (m)	UTM-y (m)	Elevation (m)	Stack height (m)	Stack temperature (K)	Stack velocity (m s^{-1})	Stack diameter (m)
TULSA RFNRY WEST	REFWEST_SN1	0.007	MONTH	228617.00	4003889.00	195.00	5.49	616.48	5.06	0.15
TULSA RFNRY WEST	REFWEST_SN2	0.005	MONTH	228750.30	4003806.26	195.10	6.71	588.71	13.20	0.15
TULSA RFNRY WEST	REFWEST_SN3	7.66	MONTH	228706.00	4002861.00	195.00	43.89	477.59	11.83	0.30
TULSA RFNRY WEST	REFWEST_SN4	0.007	MONTH	228658.38	4003859.03	195.10	7.62	547.04	7.25	0.21
TULSA RFNRY WEST	REFWEST_SN5	0.017	MONTH	228617.00	4003889.00	195.00	5.49	616.48	5.06	0.15
TULSA RFNRY WEST	REFWEST_SN6	20.44	MONTH	229176.29	4003711.77	195.10	30.48	637.59	1.92	1.62
TULSA RFNRY WEST	REFWEST_SN7	237.06	MONTH	229185.32	4003728.24	195.10	38.10	548.15	5.15	1.62
TULSA RFNRY WEST	REFWEST_SN8	41.63	MONTH	229202.04	4003723.20	195.20	18.90	505.93	2.99	1.07
TULSA RFNRY WEST	REFWEST_SN9	687.65	MONTH	228262.29	4003837.45	194.30	41.15	522.04	4.88	2.26
TULSA RFNRY WEST	REFWEST_SN10	45.53	MONTH	228237.62	4003989.27	194.20	15.24	683.15	2.99	1.37
TULSA RFNRY WEST	REFWEST_SN11	43.48	MONTH	228236.99	4003995.32	194.20	15.24	471.48	4.11	1.52
TULSA RFNRY WEST	REFWEST_SN12	0.004	MONTH	228251.07	4004028.52	193.90	25.91	768.71	4.05	1.52
TULSA RFNRY WEST	REFWEST_SN13	0.007	MONTH	228262.17	4004029.83	193.90	27.43	736.48	2.19	2.13
TULSA RFNRY WEST	REFWEST_SN14	150.00	MONTH	228246.58	4004020.78	193.90	27.74	922.04	4.82	2.13
TULSA RFNRY WEST	REFWEST_SN15	18.25	MONTH	228246.08	4004012.79	193.90	30.78	877.59	2.04	1.13
TULSA RFNRY WEST	REFWEST_SN16	65.03	MONTH	228239.18	4003982.16	194.30	23.47	523.15	26.33	0.61
TULSA RFNRY WEST	REFWEST_SN17	18.27	MONTH	229175.91	4003721.81	195.10	27.43	560.93	3.20	0.91
TULSA RFNRY WEST	REFWEST_SN18	41.40	MONTH	228239.37	4003969.12	194.60	20.12	594.26	2.38	1.37
TULSA RFNRY WEST	REFWEST_SN19	54.57	MONTH	228279.45	4003823.37	194.50	38.10	726.48	2.26	2.13
TULSA RFNRY WEST	REFWEST_SN20	210.11	MONTH	228279.45	4003823.37	194.50	38.10	738.71	4.88	2.26
TULSA RFNRY WEST	REFWEST_SN21	370.21	MONTH	228688.88	4003894.68	195.19	33.53	394.26	3.41	3.20

Table B-35. 2012 PSO Northeastern and Sapulpa point source emissions, locations, and stack parameters.

Facility Name	AERMOD source ID	Emissions (tons year^{-1})	Emission factor	UTM-x (m)	UTM-y (m)	Elevation (m)	Stack height (m)	Stack temperature (K)	Stack velocity (m s^{-1})	Stack diameter (m)
PSO NORTHEASTERN	PSO_SE1	7401.70	HOURLY	258002.59	4034618.88	195.67	182.88	394.26	13.81	8.23
PSO NORTHEASTERN	PSO_SE2	26.69	HOURLY	258002.59	4034618.88	195.67	182.88	394.26	13.81	8.23
PSO NORTHEASTERN	PSO_SE3	3.08	HOURLY	257841.41	4035283.44	195.41	55.78	393.71	16.28	5.49
PSO NORTHEASTERN	PSO_SE4	8038.60	HOURLY	258002.59	4034618.88	195.67	182.88	394.26	13.81	8.23
PSO NORTHEASTERN	PSO_SE5	19.99	HOURLY	258002.59	4034618.88	195.67	182.88	394.26	13.81	8.23
PSO NORTHEASTERN	PSO_SE6	2.27	HOURLY	257850.92	4035160.78	195.23	45.72	366.48	19.69	5.74
PSO NORTHEASTERN	PSO_SE7	2.42	HOURLY	257850.92	4035160.78	195.23	45.72	366.48	21.55	5.49
SAPULPA	SAP_SN1	100.32	MONTH	220648.04	3989373.19	215.01	28.35	530.37	9.60	1.86
SAPULPA	SAP_SN2	33.08	MONTH	220621.83	3989378.25	215.62	32.31	498.71	19.39	1.29
SAPULPA	SAP_SN3	78.85	MONTH	220621.83	3989378.25	215.62	29.87	515.37	10.27	1.71
SAPULPA	SAP_SN4	0.02	MONTH	220667.19	3989381.92	214.54	26.52	310.93	2.13	2.29
SAPULPA	SAP_SN5	0.03	MONTH	220667.19	3989381.92	214.54	29.26	310.93	2.13	2.29

Table B-36. 2012 Sapulpa area source emissions, locations, and release parameters.

Facility Name	AERMOD source ID	Emissions (tons year^{-1})	Emission factor	UTM-x (m)	UTM-y (m)	Elevation (m)	Release height (m)	X-dimension (m)	Y-dimension (m)	Angle	σ_z (m)
SAPULPA	SAP_SN6	0.03	MONTH	220691.84	3989080	218.06	10.67	2.74	2.74	0	2.48

Table B-37. 2013 Tulsa East Refinery point source emissions, locations, and stack parameters.

Facility Name	AERMOD source ID	Emissions (tons year^{-1})	Emission factor	UTM-x (m)	UTM-y (m)	Elevation (m)	Stack height (m)	Stack temperature (K)	Stack velocity (m s^{-1})	Stack diameter (m)
TULSA RFNRY-EAST	REFEAST_SN1	11.85	MONTH	230409.02	4000701.87	192.12	73.15	1088.71	43.34	0.49
TULSA RFNRY-EAST	REFEAST_SN2	0.34	MONTH	229761.77	4000607.68	192.00	30.78	317.59	6.68	0.76
TULSA RFNRY-EAST	REFEAST_SN4	0.26	MONTH	229823.09	4000610.90	192.00	60.96	444.26	5.88	0.61
TULSA RFNRY-EAST	REFEAST_SN5	5.08	MONTH	229944.74	4000860.87	194.00	58.22	572.59	22.92	1.52
TULSA RFNRY-EAST	REFEAST_SN6	0.10	MONTH	229658.38	4000653.14	192.00	29.26	313.71	8.23	1.13
TULSA RFNRY-EAST	REFEAST_SN7	0.10	MONTH	229663.74	4000658.82	192.00	30.48	311.48	7.86	1.13
TULSA RFNRY-EAST	REFEAST_SN8	0.05	MONTH	229954.38	4001000.54	192.90	13.72	570.37	16.06	1.07
TULSA RFNRY-EAST	REFEAST_SN9	0.22	MONTH	229946.71	4000617.28	194.83	42.67	583.15	14.60	1.46
TULSA RFNRY-EAST	REFEAST_SN11	0.45	MONTH	229945.36	4000870.85	194.00	46.02	624.82	3.99	1.77
TULSA RFNRY-EAST	REFEAST_SN12	0.83	MONTH	229956.32	4001096.60	194.47	53.34	449.82	3.47	3.51
TULSA RFNRY-EAST	REFEAST_SN14	0.44	MONTH	229971.12	4000687.91	194.77	38.10	466.48	3.84	2.53
TULSA RFNRY-EAST	REFEAST_SN15	0.57	MONTH	229950.17	4000673.17	194.97	37.80	560.93	10.00	1.77
TULSA RFNRY-EAST	REFEAST_SN16	0.11	MONTH	229950.84	4000700.18	195.00	37.80	533.15	7.04	1.37
TULSA RFNRY-EAST	REFEAST_SN17	0.99	MONTH	229912.85	4001441.17	192.39	21.64	449.82	6.25	2.13
TULSA RFNRY-EAST	REFEAST_SN18	1.02	MONTH	229940.84	4001441.17	192.31	21.64	449.82	6.19	2.13

Table B-38. 2013 Tulsa West Refinery point source emissions, locations, and stack parameters.

Facility Name	AERMOD source ID	Emissions (tons $year^{-1}$)	Emission factor	UTM-x (m)	UTM-y (m)	Elevation (m)	Stack height (m)	Stack temperature (K)	Stack velocity (m s^{-1})	Stack diameter (m)
TULSA RFNRY WEST	REFWEST_SN1	8.22	MONTH	228706.00	4002861.00	195.00	43.89	477.59	11.83	0.30
TULSA RFNRY WEST	REFWEST_SN2	0.15	MONTH	228659.61	4003895.03	195.10	18.29	433.15	8.23	1.52
TULSA RFNRY WEST	REFWEST_SN3	0.26	MONTH	228660.10	4003903.01	195.10	18.29	440.37	6.49	1.52
TULSA RFNRY WEST	REFWEST_SN4	0.10	MONTH	228658.38	4003859.03	195.10	24.38	425.93	6.25	1.52
TULSA RFNRY WEST	REFWEST_SN5	0.02	MONTH	228617.00	4003889.00	195.00	5.49	616.48	5.06	0.15
TULSA RFNRY WEST	REFWEST_SN6	9.09	MONTH	229176.29	4003711.77	195.10	30.48	637.59	1.92	1.62
TULSA RFNRY WEST	REFWEST_SN7	169.39	MONTH	229185.32	4003728.24	195.10	38.10	548.15	5.15	1.62
TULSA RFNRY WEST	REFWEST_SN8	26.45	MONTH	229202.04	4003723.20	195.20	18.90	505.93	2.99	1.07
TULSA RFNRY WEST	REFWEST_SN9	360.29	MONTH	228262.29	4003837.45	194.30	41.15	522.04	4.88	2.26
TULSA RFNRY WEST	REFWEST_SN10	8.45	MONTH	228237.62	4003989.27	194.20	15.24	683.15	2.99	1.37
TULSA RFNRY WEST	REFWEST_SN11	16.96	MONTH	228236.99	4003995.32	194.20	15.24	471.48	4.11	1.52
TULSA RFNRY WEST	REFWEST_SN12	0.002	MONTH	228251.07	4004028.52	193.90	25.91	768.71	4.05	1.52
TULSA RFNRY WEST	REFWEST_SN13	0.003	MONTH	228262.17	4004029.83	193.90	27.43	736.48	2.19	2.13
TULSA RFNRY WEST	REFWEST_SN14	36.95	MONTH	228246.58	4004020.78	193.90	27.74	922.04	4.82	2.13
TULSA RFNRY WEST	REFWEST_SN15	4.42	MONTH	228246.08	4004012.79	193.90	30.78	877.59	2.04	1.13
TULSA RFNRY WEST	REFWEST_SN16	23.79	MONTH	228239.18	4003982.16	194.30	23.47	523.15	26.33	0.61
TULSA RFNRY WEST	REFWEST_SN17	10.56	MONTH	229175.91	4003721.81	195.10	27.43	560.93	3.20	0.91
TULSA RFNRY WEST	REFWEST_SN18	10.76	MONTH	228239.37	4003969.12	194.60	20.12	594.26	2.38	1.37
TULSA RFNRY WEST	REFWEST_SN19	34.20	MONTH	228279.45	4003823.37	194.50	38.10	726.48	2.26	2.13
TULSA RFNRY WEST	REFWEST_SN20	124.53	MONTH	228279.45	4003823.37	194.50	38.10	738.71	4.88	2.26
TULSA RFNRY WEST	REFWEST_SN21	0.03	MONTH	228524.37	4004105.79	195.40	27.74	555.37	3.02	1.22

Table B-39. 2013 Tulsa West Refinery point source emissions, locations, and stack parameters.

Facility Name	AERMOD source ID	Emissions (tons year^{-1})	Emission factor	UTM-x (m)	UTM-y (m)	Elevation (m)	Stack height (m)	Stack temperature (K)	Stack velocity (m s^{-1})	Stack diameter (m)
TULSA RFNRY WEST	REFWEST_SN22	0.14	MONTH	229194.24	4003726.69	195.20	34.14	478.71	3.20	1.83
TULSA RFNRY WEST	REFWEST_SN23	0.07	MONTH	228527.85	4004113.59	195.10	34.14	610.93	2.47	1.68
TULSA RFNRY WEST	REFWEST_SN24	211.21	MONTH	228688.88	4003894.68	195.19	33.53	394.26	3.41	3.20

Table B-40. 2013 PSO Northeastern and Sapulpa point source emissions, locations, and stack parameters.

Facility Name	AERMOD source ID	Emissions (tons year^{-1})	Emission factor	UTM-x (m)	UTM-y (m)	Elevation (m)	Stack height (m)	Stack temperature (K)	Stack velocity (m s^{-1})	Stack diameter (m)
PSO NORTHEASTERN	PSO_SE1	9337.20	HOURLY	258002.59	4034618.88	195.67	182.88	394.26	13.81	8.23
PSO NORTHEASTERN	PSO_SE2	22.32	HOURLY	258002.59	4034618.88	195.67	182.88	394.26	13.81	8.23
PSO NORTHEASTERN	PSO_SE3	1.38	HOURLY	257841.41	4035283.44	195.41	55.78	393.71	16.28	5.49
PSO NORTHEASTERN	PSO_SE4	9007.50	HOURLY	258002.59	4034618.88	195.67	182.88	394.26	13.81	8.23
PSO NORTHEASTERN	PSO_SE5	38.16	HOURLY	258002.59	4034618.88	195.67	182.88	394.26	13.81	8.23
PSO NORTHEASTERN	PSO_SE6	2.88	HOURLY	257850.92	4035160.78	195.23	45.72	366.48	19.69	5.74
PSO NORTHEASTERN	PSO_SE7	3.11	HOURLY	257850.92	4035160.78	195.23	45.72	366.48	21.55	5.49
SAPULPA	SAP_SN1	0.01	MONTH	220685.88	3989163.75	216.57	2.44	755.37	21.73	0.10
SAPULPA	SAP_SN2	108.29	MONTH	220648.04	3989373.19	215.01	28.35	530.37	9.60	1.86
SAPULPA	SAP_SN3	33.74	MONTH	220621.83	3989378.25	215.62	32.31	498.71	19.39	1.29
SAPULPA	SAP_SN4	98.48	MONTH	220621.83	3989378.25	215.62	29.87	515.37	10.27	1.71
SAPULPA	SAP_SN5	0.02	MONTH	220667.19	3989381.92	214.54	26.52	310.93	2.13	2.29
SAPULPA	SAP_SN6	0.03	MONTH	220667.19	3989381.92	214.54	29.26	310.93	2.13	2.29

Table B-41. 2013 Sapulpa area source emissions, locations, and release parameters.

Facility Name	AERMOD source ID	Emissions (tons $year^{-1}$)	Emission factor	UTM-x (m)	UTM-y (m)	Elevation (m)	Release height (m)	X-dimension (m)	Y-dimension (m)	Angle	σ_z (m)
SAPULPA	SAP_SN7	0.03	MONTH	220691.84	3989080	218.06	10.67	2.74	2.74	0	2.48

REFERENCES

UNC (University of North Carolina). (2017). Sparse Matrix Operator Kernel Emissions modeling system User Manual. Available at: https://www.cmascenter.org/help/documentation.cfm?MODEL=smoke&VERSION=4.5

U.S. EPA. (2016). User's Guide for the AMS/EPA Regulatory Model – AERMOD. EPA-454/B-16-011. U.S. Environmental Protection Agency, Research Triangle Park, NC 27711.

APPENDIX C

AIR QUALITY MODELING DOMAINS FOR STUDY AREAS

Preface: The modeling domains, including receptors and modeled sources, for the three study areas are shown in Figures C-1 and C-2, for Fall River, Figures C-3 and C-4 for Indianapolis, and Figures C-5 and C-6 for Tulsa. Sources are denoted by stars, monitors by triangles, and gridded receptors by small dots. The blue airport symbol denotes the location of the NWS station used in the modeling.

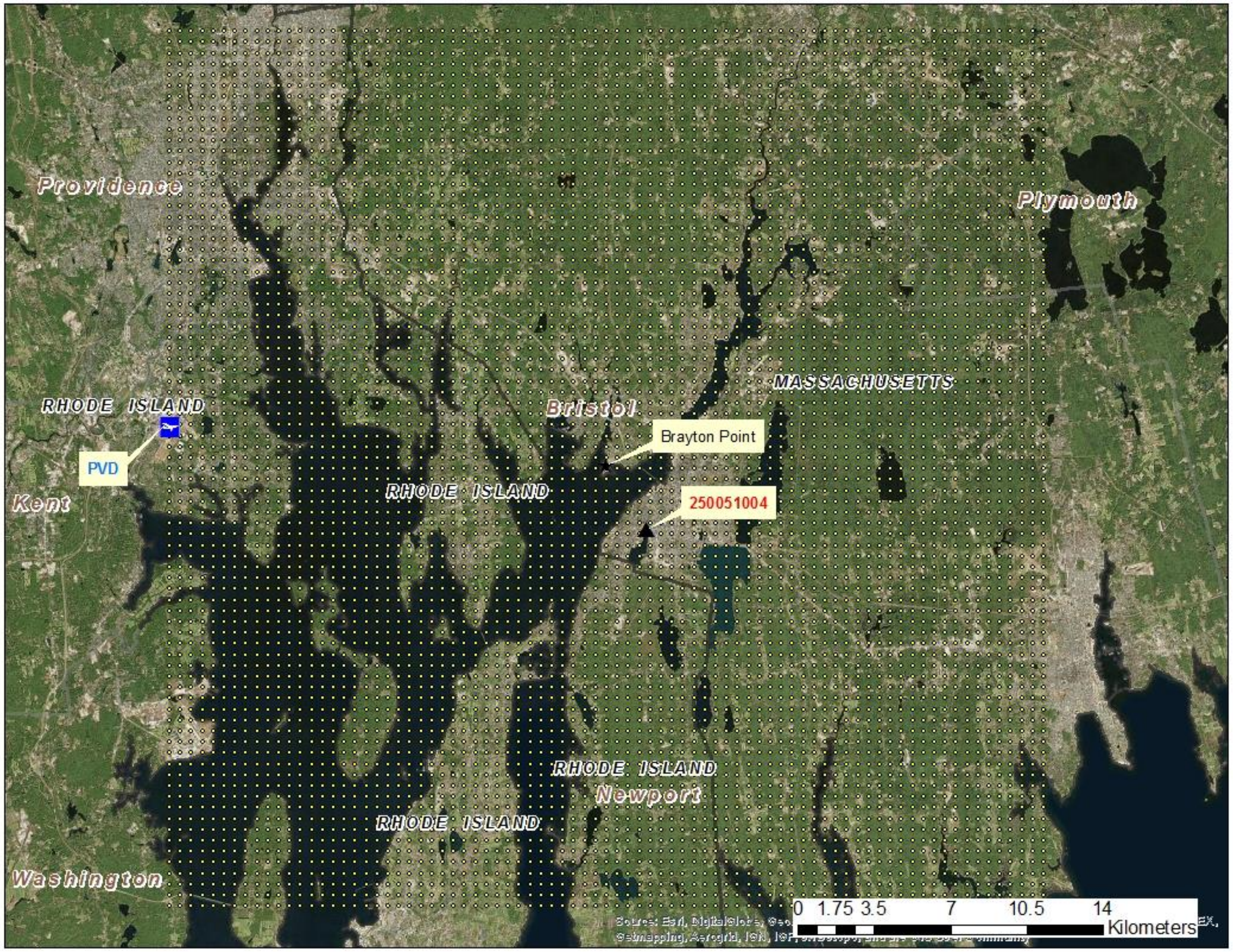


Figure C-1. Fall River study area air quality modeling domain.

Figure C-2. Detailed view of Fall River study area air quality modeling domain.

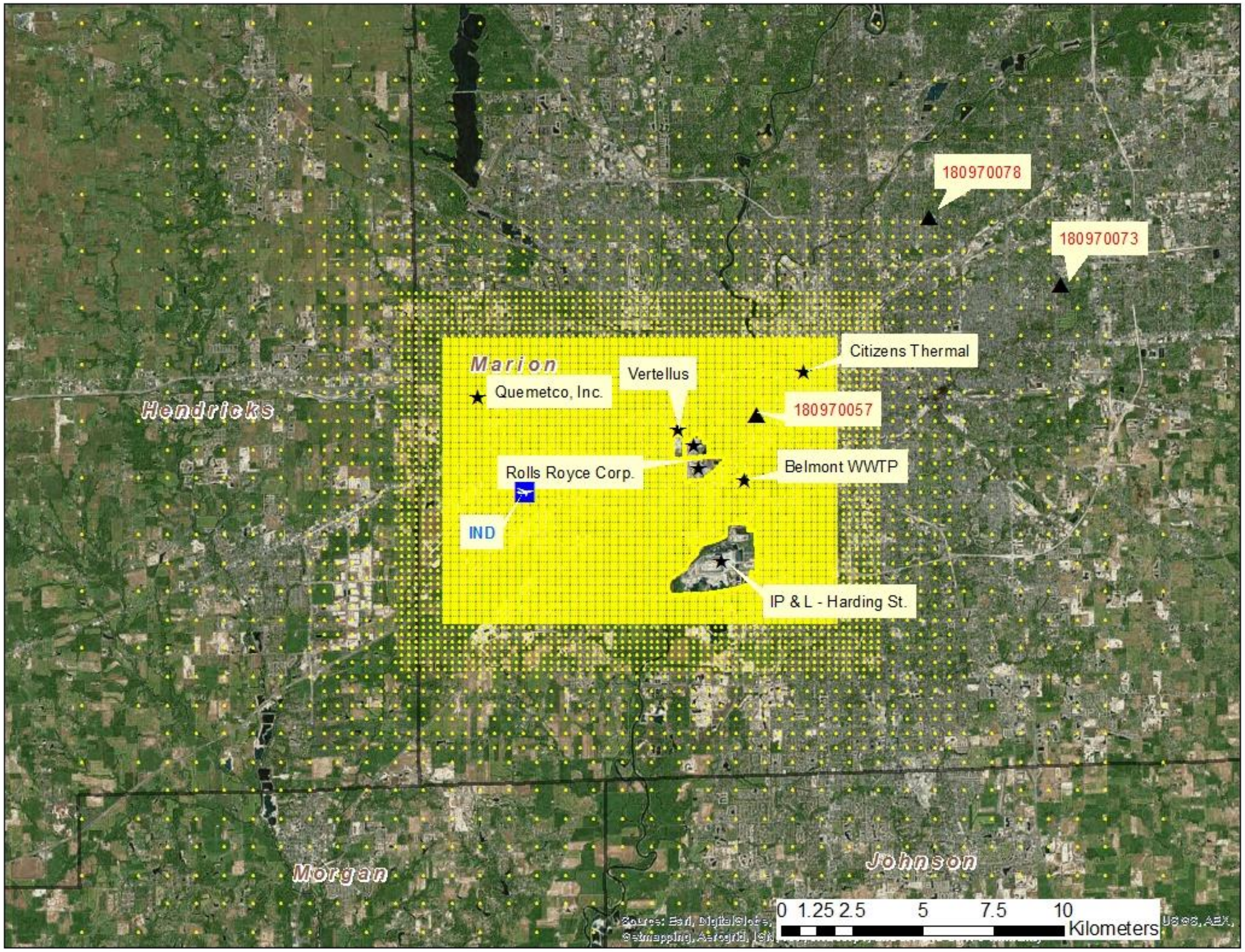


Figure C-3. Indianapolis study area air quality modeling domain.

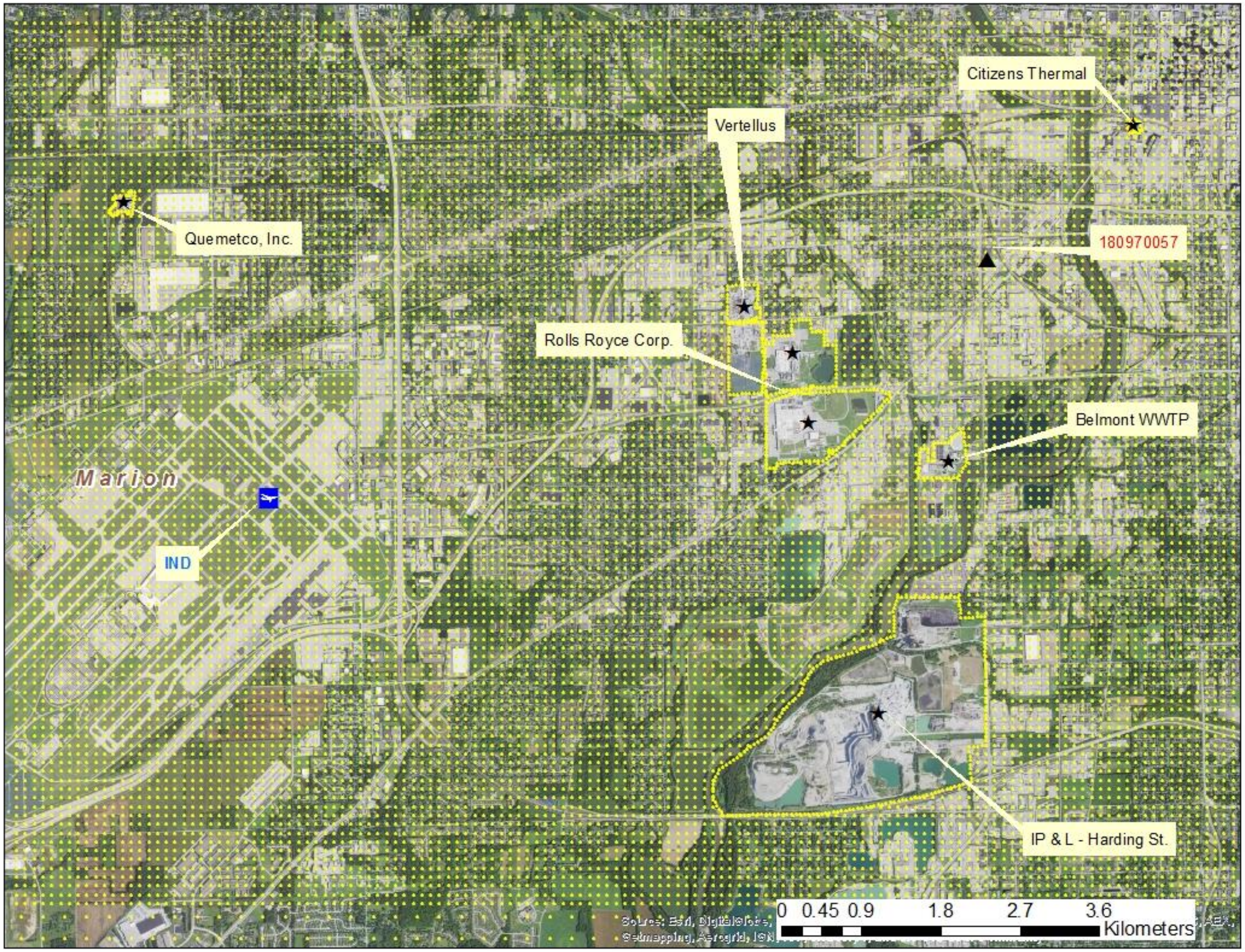


Figure C-4. Detailed view of Indianapolis study area air quality modeling domain.

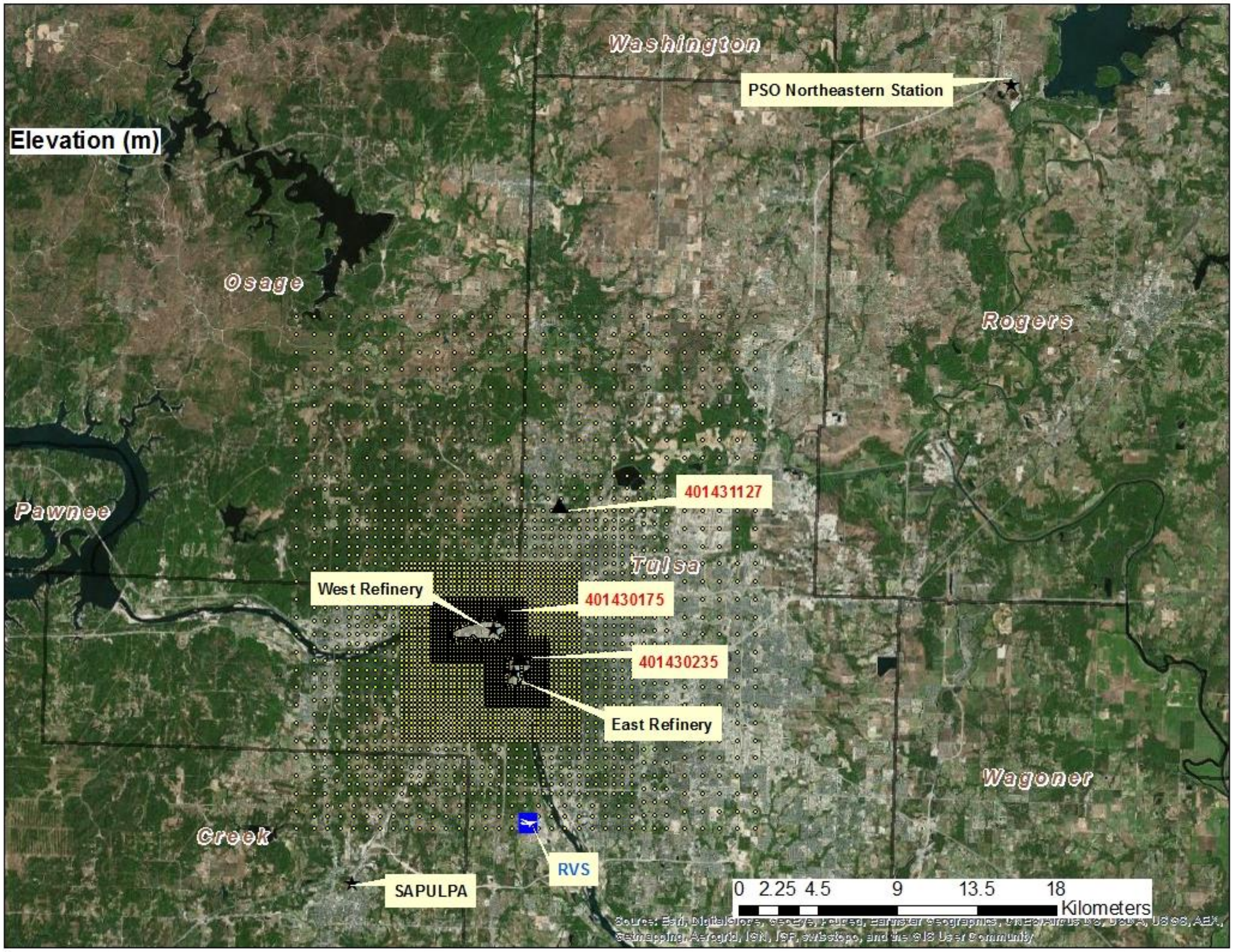


Figure C-5. Tulsa study area air quality modeling domain.

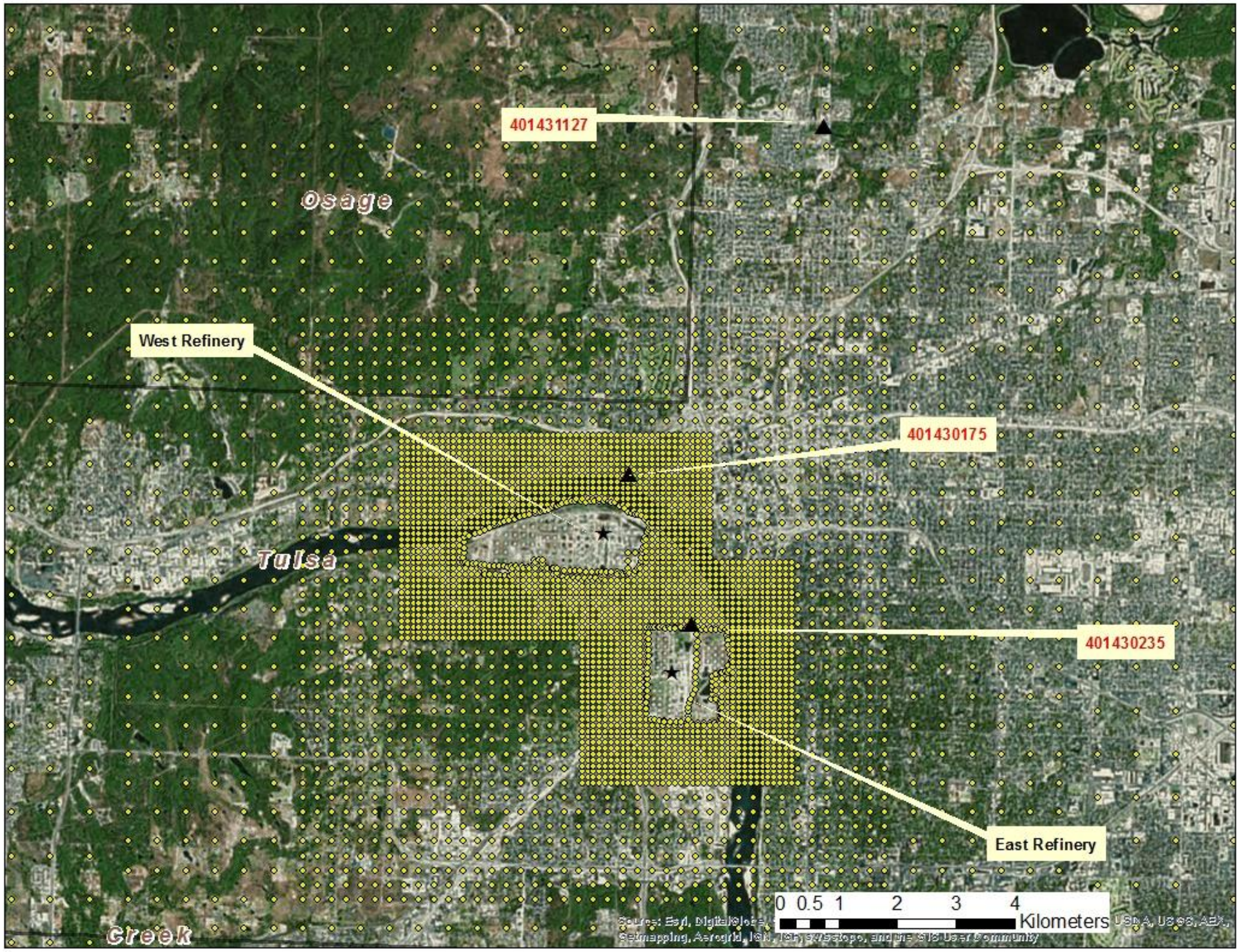


Figure C-6. Detailed view of Tulsa study area air quality modeling domain.

APPENDIX D

MODELED AIR QUALITY EVALUATION

AERMOD output for the three study areas was evaluated using three methods. First, comparison of the 99th percentile of daily 1-hour maximum concentrations for each and subsequent 3-year design values were compared at each monitor. Second, simple QQ-plots were generated to provide a quick visual performance of the model for 1-hour, 3-hour, and 24-hour averages. The QQ-plots are comparisons of the observed and modeled concentrations, unpaired in time and space, consistent with regulatory evaluations of AERMOD (U.S. EPA, 2003; Venkatram et al., 2001). Third, for a more rigorous comparison, the EPA Protocol for determining best performing model, or sometimes called the Cox-Tikvart method (U.S. EPA, 1992; Cox and Tikvart, 1990) was used. Normally, this protocol is used to determine which model or model scenarios among a suite of models or scenario is the better performer for regulatory application and focuses on the higher concentrations in the concentration distribution as these are the concentrations of interest in most regulatory applications (State Implementation Plans and Prevention of Significant Deterioration). For example, U.S. EPA (2016) used the protocol to determine which was a better performer in terms of meteorological data, observed or prognostic data. For the study presented here, we are only evaluating one model and one scenario, i.e., AERMOD for 2011-2013. Therefore, the protocol will not be used to its full extent, but rather to provide information regarding the performance of the model for these study areas. An explanation of the protocol follows.

The protocol uses fractional bias (equation D-1) for evaluating model performance.

$$FB = 2\left[\frac{OB-PR}{OB+PR}\right] \qquad \text{Equation D-1}$$

Where FB is the fractional bias, OB is the average of the highest 25 observed concentrations and PR is the average of the highest 25 predicted averages.

In the evaluation, air quality models are subjected to a comprehensive statistical comparison that involves both an operational and scientific component. The operational component is to measure the model’s ability to estimate concentration statistics most directly used for regulatory purposes and the scientific component evaluates the model’s ability to perform accurately throughout the range of meteorological conditions and the geographic area of concern (U.S. EPA, 1992). The test statistic used for the comparison is the robust highest concentration (RHC) statistic and is given by:

$$RHC = \mathrm{X}(N) + [\bar{\mathrm{X}} - \mathrm{X}(N)] \times \ln\left[\frac{3N-1}{2}\right] \qquad \text{Equation D-2}$$

Where *X(N)* is the Nth largest value, $\overline{X}$ is the average of *N-1* values, and *N* is the number of values exceeding the threshold value, usually 26.

The operational component of the evaluation compares performance in terms of the largest network-wide RHC test statistic. The RHC is calculated separately for each monitor within the network for both observed and modeled values. The absolute fractional bias (AFB) is calculated for both 3 and 24-hour averages using the absolute value of the results of equation 1. The inputs to the AFB calculation are the highest observed RHC and the highest modeled RHC.

The scientific component of the evaluation is also based on absolute fractional bias but the bias is calculated using the RHC for each meteorological condition and monitor. The meteorological conditions are a function of atmospheric stability and wind speed. For the purposes of these studies, six unique conditions were defined based on two wind speed categories (below and above 2.0 m/s) and three stability categories: unstable, neutral, and stable. [1] In scientific evaluation, only 1-hour concentrations are used and the AFB is based on RHC values paired in space and stability/wind speed combination.

A composite performance measure (CPM) is calculated from the 1-hour, 3-hour, and 24-hour AFB's:

$$CPM = \frac{1}{3} \times \left(\overline{AFB_{i,j}}\right) + \frac{2}{3} \times \left[\frac{AFB_3 - AFB_{24}}{2}\right] \qquad \text{Equation D-3}$$

Where $AFB_{i,j}$ is the absolute fractional bias for monitor i and meteorological condition j, $\overline{AFB_{i,j}}$ is the average absolute fractional bias across all monitors and meteorological conditions, AFB_3 is the absolute fractional bias for the 3-hour average, and AFB_{24} is the absolute fractional bias for the 24-hour average. The closer the CPM is to zero, the better the performance of the model. Also, since the absolute fraction biases are calculated using equation 1, which is bounded by 2 (U.S. EPA, 1992), then the maximum value for the CPM is also 2.

Both the QQ-plots and the EPA protocol are applied to the model output in two ways. First, evaluations were conducted by comparing model output and observations unpaired in time and space, consistent with regulatory evaluations of AERMOD (U.S. EPA, 2003; Venkatram et al., 2001). In regulatory applications, the emphasis is not on where potential modeled NAAQS violations occur, but whether they occur. Second, given the nature of this particular study as an exposure analysis, where individual receptors are being used on an hourly basis, the QQ-plots and the EPA protocol were both applied to model output at individual monitors. This would be a pairing in space but not necessarily time. This would help answer the question, is the model

[1] In U.S. EPA (1992), the three stability categories are related to the Pasquill-Gifford categories, unstable being A, B, and C, neutral being D, and stable being E and F. Since AERMOD does not use the stability categories, the stability class was determined using Monin-Obukhov length and surface roughness using methodology from AERMOD subroutine LTOPG.

performing well at predicting the locations of concentrations of interest. Also, since the monitors in each of the study areas are located near populations, if the model performs well near these monitors then reasonable performance in the population areas, or areas of interest for exposure, can be expected. For all three areas, QQ-plots and the EPA protocol were performed for the entire three-year period, 2011-2013, and for each year individually to see if individual years were driving the total period comparisons.

Fall River: Modeled Air Quality Evaluation

Only one monitor (Figure C-1, Figure C-2) was located in the vicinity of Brayton Power Station. Table D-1 shows the monitored and modeled annual 99th percentile daily 1-hour maximum concentration and the three-year design value. With the exception of 2011, the model under-predicts the 99th percentile of the daily 1-hour maximum concentration and under-predicts the 3-year design value.

Table D-1. Fall River monitored and modeled annual 99th percentile daily 1-hour maximum concentrations (μg m^{-3}) and 3-year design value (μg m^{-3}).

Year	Monitor	Observed	Model
2011	250051004	169.8	177.1
2012	250051004	171.1	138.2
2013	250051004	161.9	84.9
Design Value	250051004	167.6	133.4

Figures D-1 through D-3 show the QQ-plots for 1-hour, 3-hour, and 24-hour averages respectively. In each figure, panel a is the ranked comparisons for the entire 3-year period, while panels b-d are the individual years' ranked pairings. For the 1-hour comparison across all three years, the model is over predicting at the lower end of the concentration distributions (less than 50 μg m^{-3}), predicts very well at the middle of the distribution (50 -125 μg m^{-3}) and then shifts to under-prediction from 150 to 250 μg m^{-3}. At the very high end, i.e. the last three observations, the model over-predicts, under-predicts and is almost equal to the highest monitored concentration. Analyzing the three individual years, the model appears to perform the best in 2011. The 3-hour QQ-plots exhibit similar patterns as the 1-hour plots. The 24-hour plots exhibit a pattern of over-prediction at the low to mid-range of the distributions and then under prediction at the high ends.

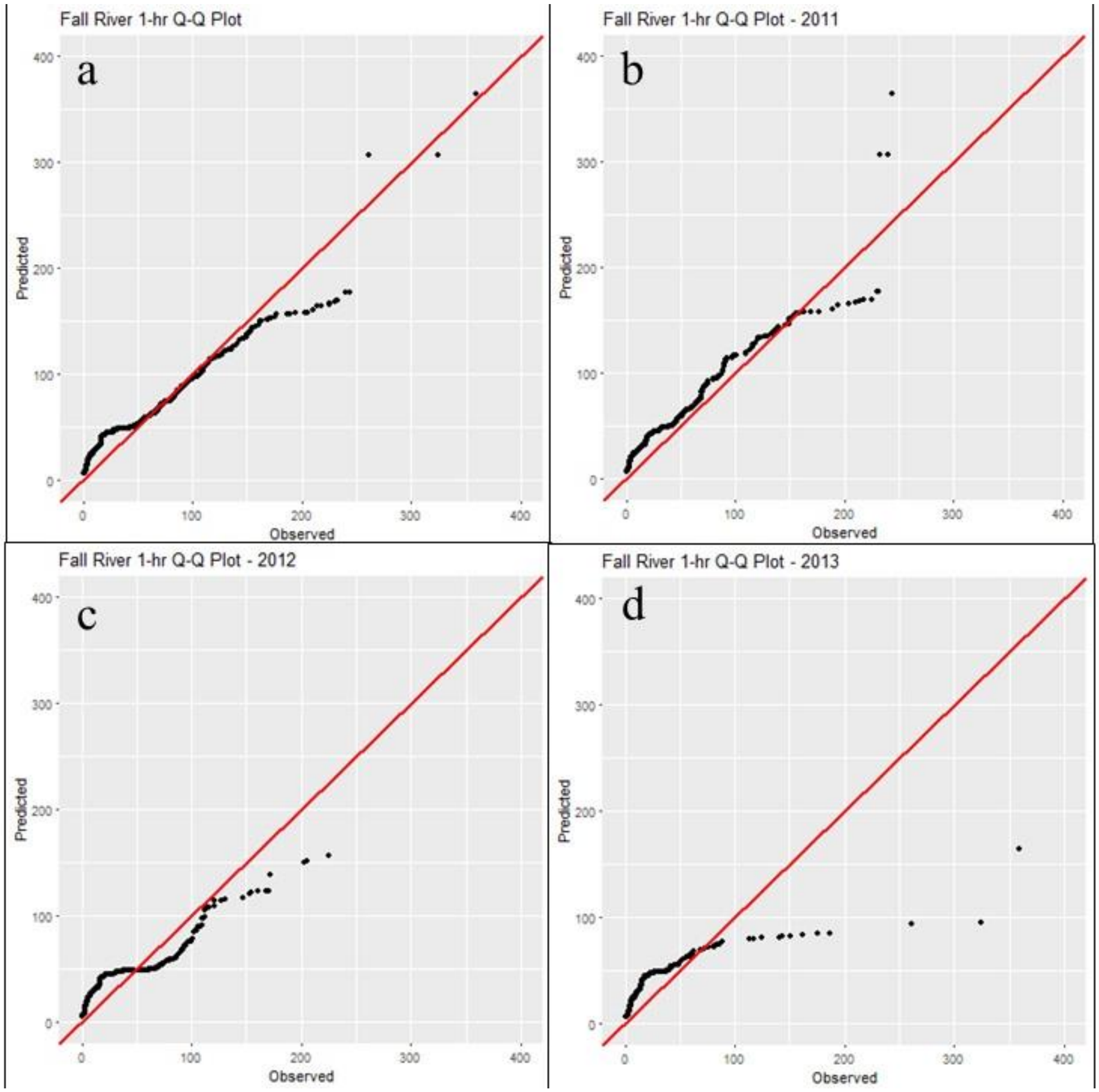


Figure D-1. Fall River 1-hour QQ plots.

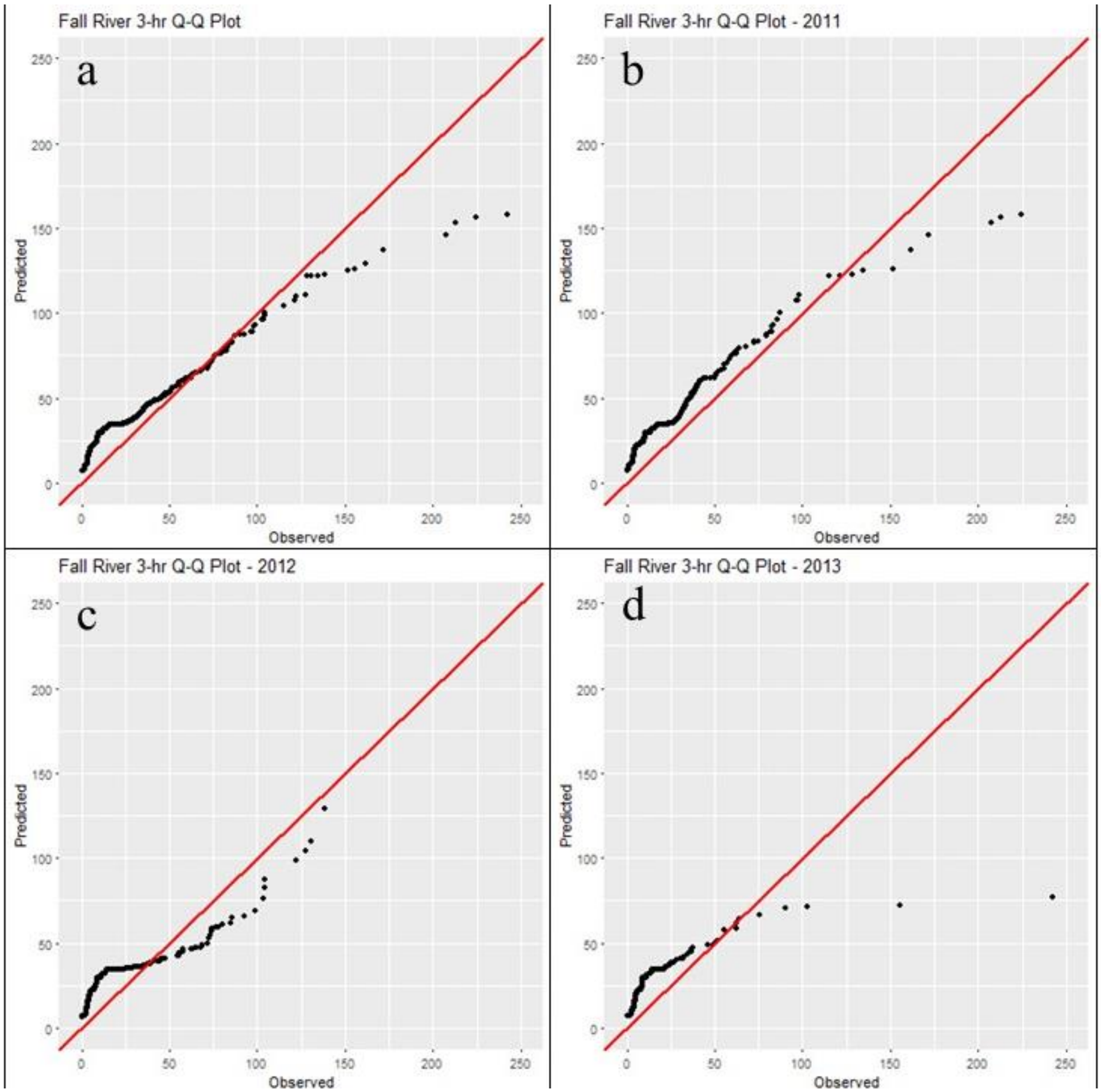


Figure D-2. Fall River 3-hour QQ plots.

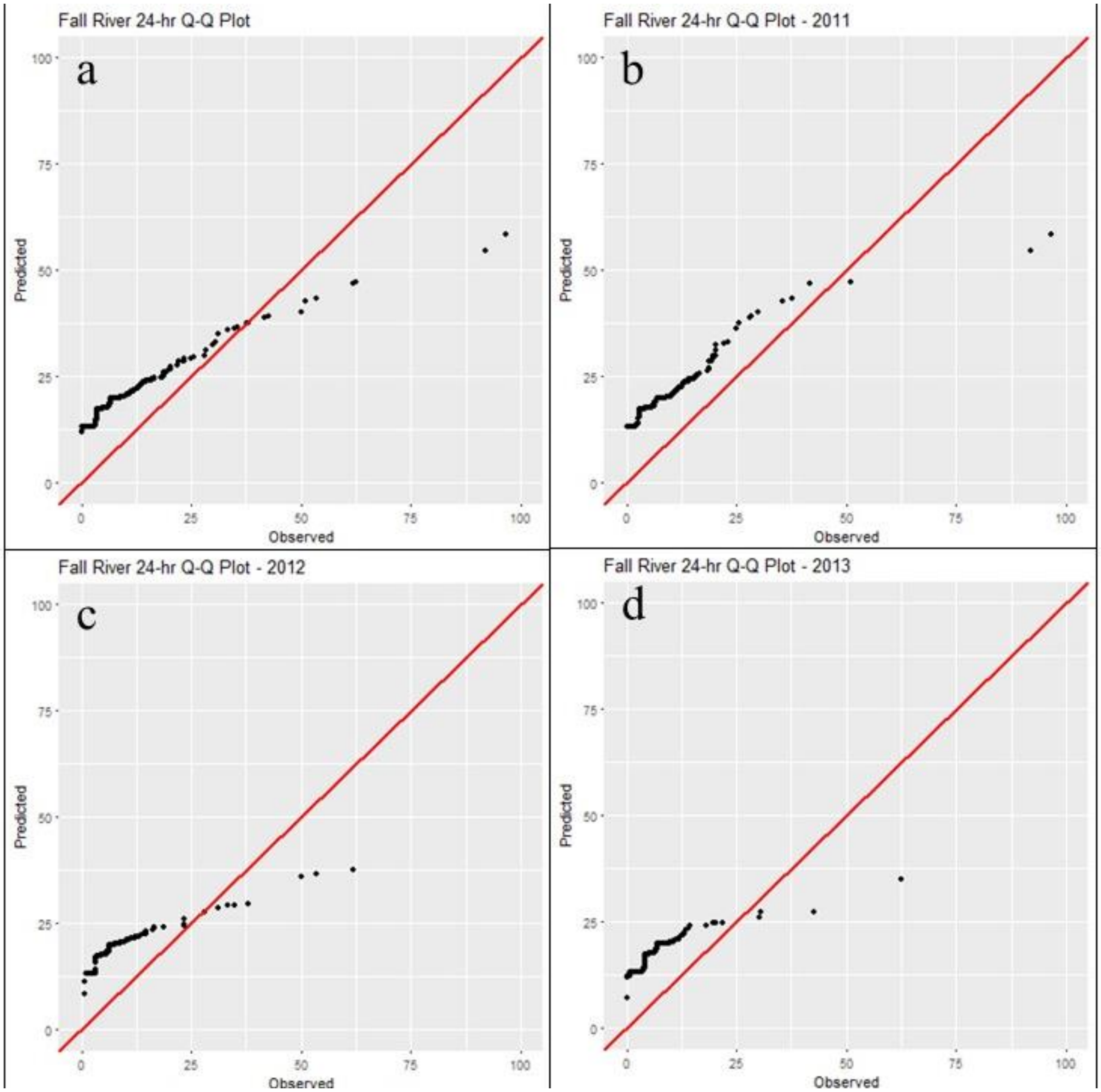


Figure D-3. Fall River QQ-plots.

In addition to the QQ-plots, composite performance metrics, CPM, were calculated for the entire period and each of the individual years.

Table D-2 lists the CPM values for 2011-13 and CPM values for the individual years. Also shown are the absolute fractional biases for 1-hour, 3-hour, and 24-hours. Overall, considering impacts from the three averaging periods, 2011 was the better performing year of the three years and the 2011-2013 CPM shows the influence of 2013.

Table D-2. Fall River composite performance metrics (CPM) and absolute fractional biases for 1-hour, 3-hour, and 24-hour averages.

Period	CPM	$AFB_{1\text{-hr}}$	$AFB_{3\text{-hr}}$	$AFB_{24\text{-hr}}$
2011-2013	0.45	0.68	0.30	0.38
2011	0.29	0.56	0.21	0.10
2012	0.35	0.43	0.22	0.41
2013	0.49	0.75	0.52	0.20

Indianapolis: Modeled Air Quality Evaluation

Three monitors were available for model evaluation in Indianapolis (Figure C-3). Table D-3 lists the annual 99th percentile daily 1-hour maximum concentration and 3-year design value for each monitor. The model is over-predicting at monitor 180970057 (the nearest monitor to the sources) and generally under-predicting each year and the design values at the other monitors. The modeled design value at 180970057 is within 10% of the monitored design value while at the other monitors, the modeled design values are within 3% of the monitored design values.

Table D-3. Indianapolis monitored and modeled annual 99th percentile daily 1-hour maximum concentrations ($\mu g\ m^{-3}$) and 3-year design value ($\mu g\ m^{-3}$).

Monitor	Year	Observed	Modeled
180970057	2011	164.8	268.7
	2012	239.4	330.8
	2013	204.3	367.5
	Design Value	202.8	322.4
180970073	2011	155.6	122.1
	2012	146.8	129.9
	2013	110.7	151.4
	Design Value	137.7	134.4
180970078	2011	156.2	153.9
	2012	159.9	162.2
	2013	182.4	168.7
	Design Value	166.1	161.6

One-hour, 3-hour, and 24-hour QQ-plots across all three monitors are shown in Figures D-4 through D-6, respectively. For 1-hour averages, the 3-year QQ-plot and 2012 and 2013 QQ-plots show an over-prediction trend except at the higher concentrations for the 3-year period and 2012, where there is under-prediction. Analysis of the 2012 higher 1-hour concentrations (Figure D-4) showed very high observations for those years which the model did not simulate while 2011 actually shows very good model performance. For the 3-hour averages (Figure D-5), all three years and the entire period show good model to monitor agreement with some over-prediction in 2013. The 24-hour averages (Figure D-6), the 3-year period and each individual year exhibit over-prediction. Overall, 2011 appeared to show the better performance among the years among all averaging periods. Figures D-7 through D-9 show the 1-hour, 3-hour, and 24-

hour QQ-plots for the individual monitors for the 3-year period and by year. Results were mixed among the three monitors. For the 1-hour averages, monitor 180970057, the closest monitor to the modeled sources (Figure C-3, Figure C-4), the modeled concentrations were higher than monitored values except at the highest concentrations for 2012. The annual 99th percentile daily 1-hour maxima and design value in Table D-3 reflect the over-prediction. For 2013, the model overestimated throughout the distribution. For the other two monitors, the modeled values showed good agreement through most of the concentration distribution and then tended toward underestimation at the higher end of the distributions. The same general trend was seen with the 3-hour average concentrations (Figure D-8) for monitor 180970073. All three monitors exhibit over-prediction for the 24-hour period (Figure D-9) which could be a consequence of the seasonal background included in the Indianapolis modeling as 3-year average background concentrations are added to each individual hour in the modeling. The inclusion of an average background could be over-estimated for some individual hours and when calculating a multi-hour average, e.g. 3 or 24-hour, the overestimates could accumulate over time.

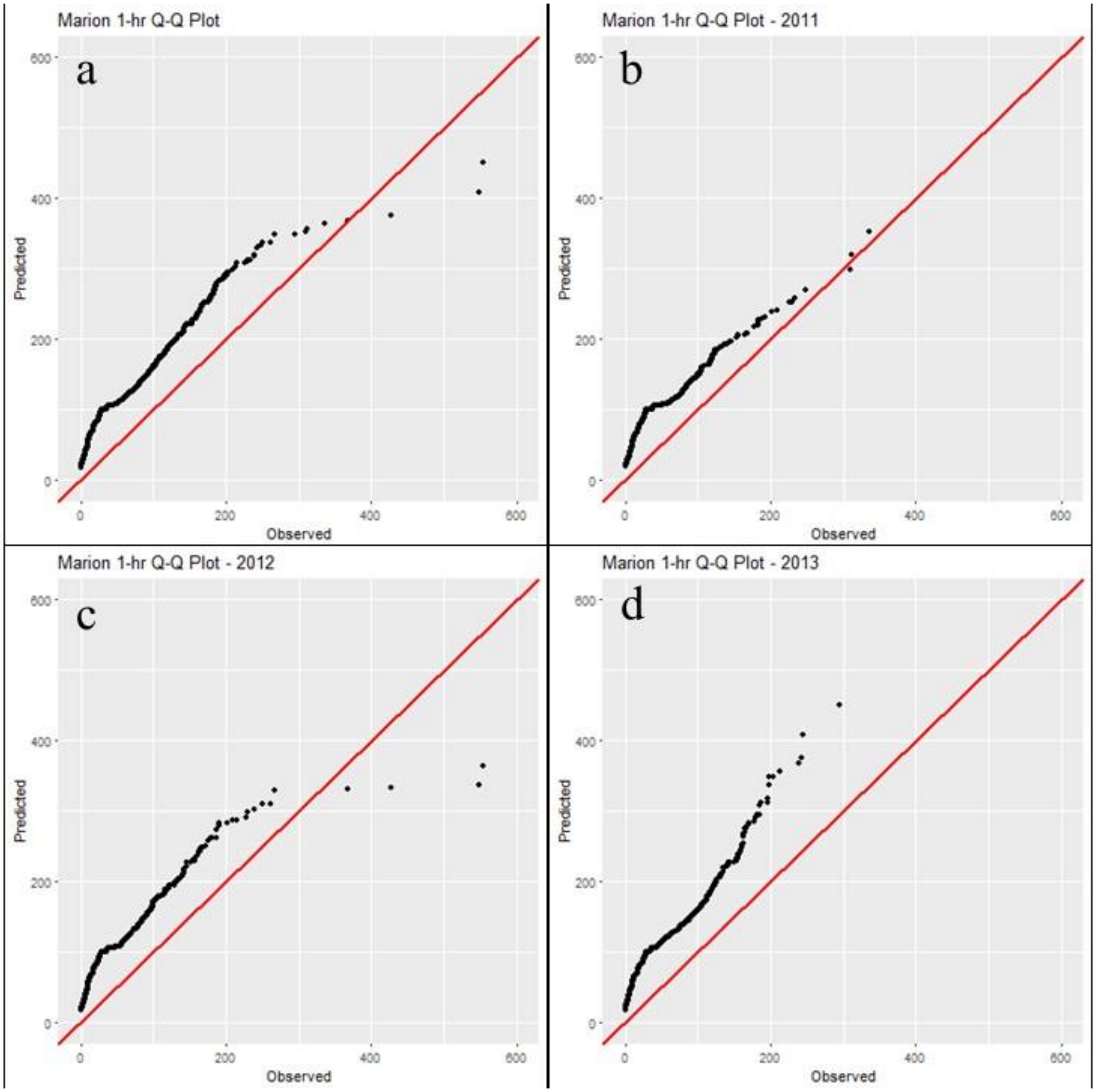


Figure D-4. Indianapolis 1-hour QQ-plots.

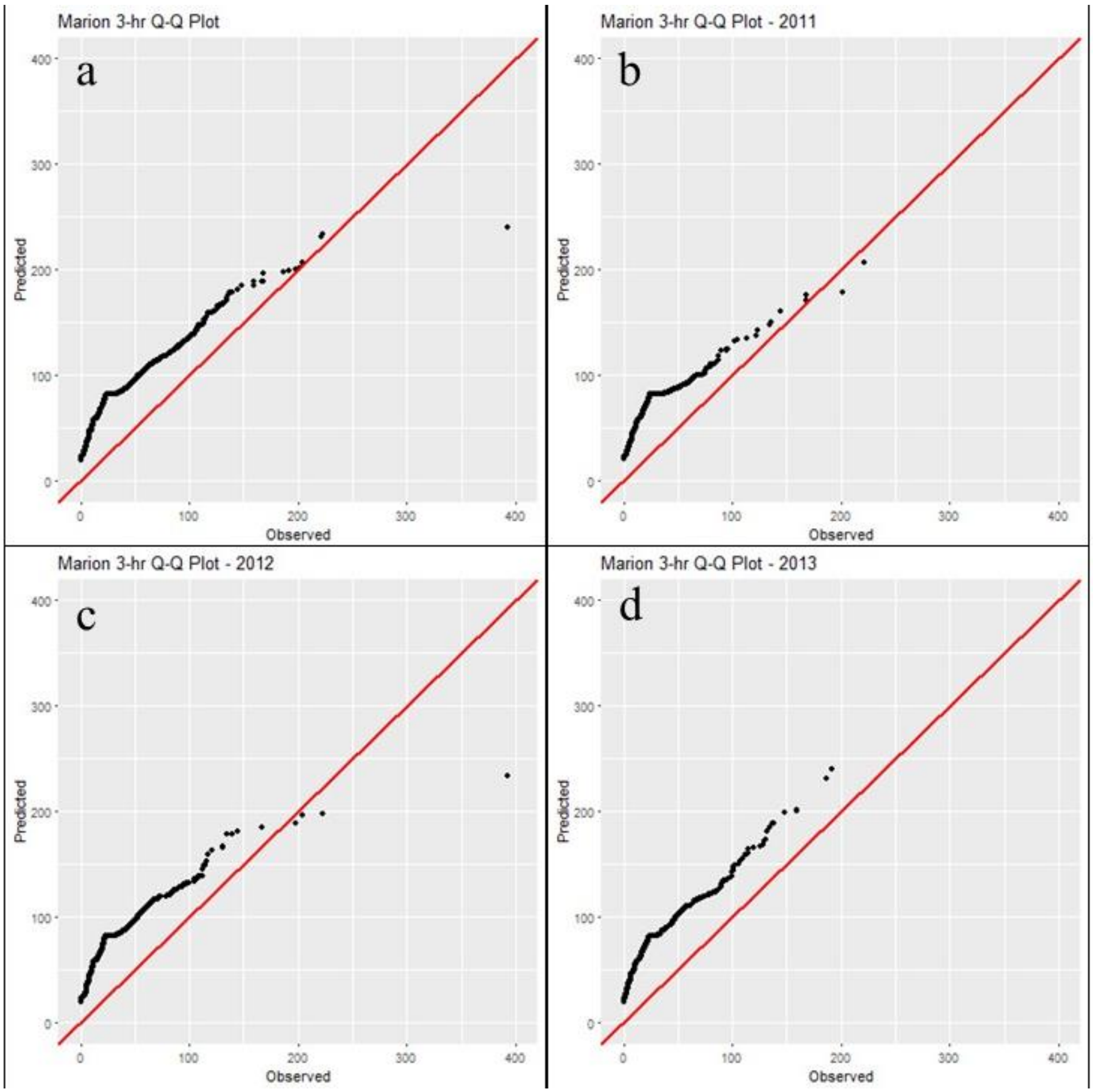


Figure D-5. Indianapolis 3-hour QQ-plots.

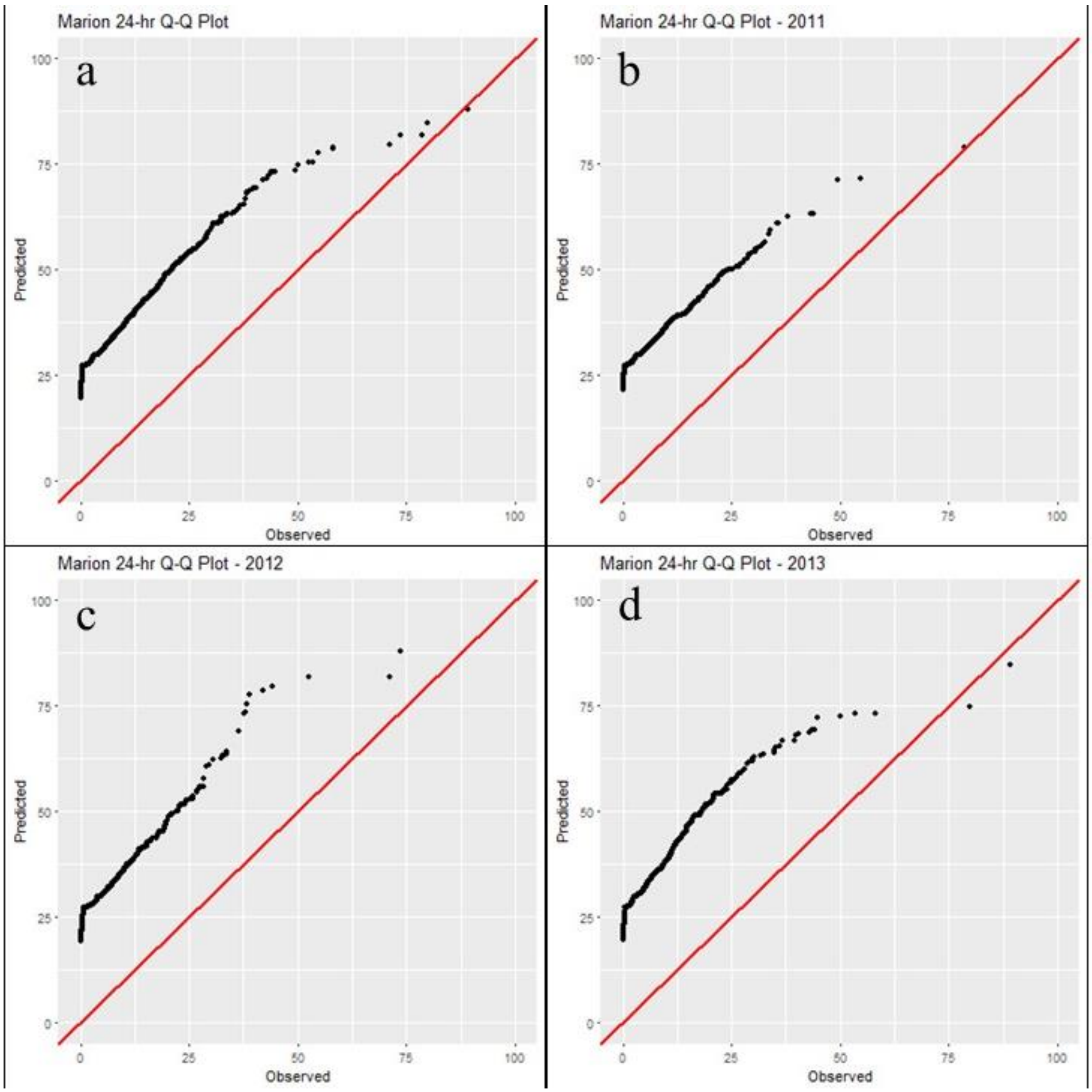


Figure D-6. Indianapolis 24-hour QQ-plots.

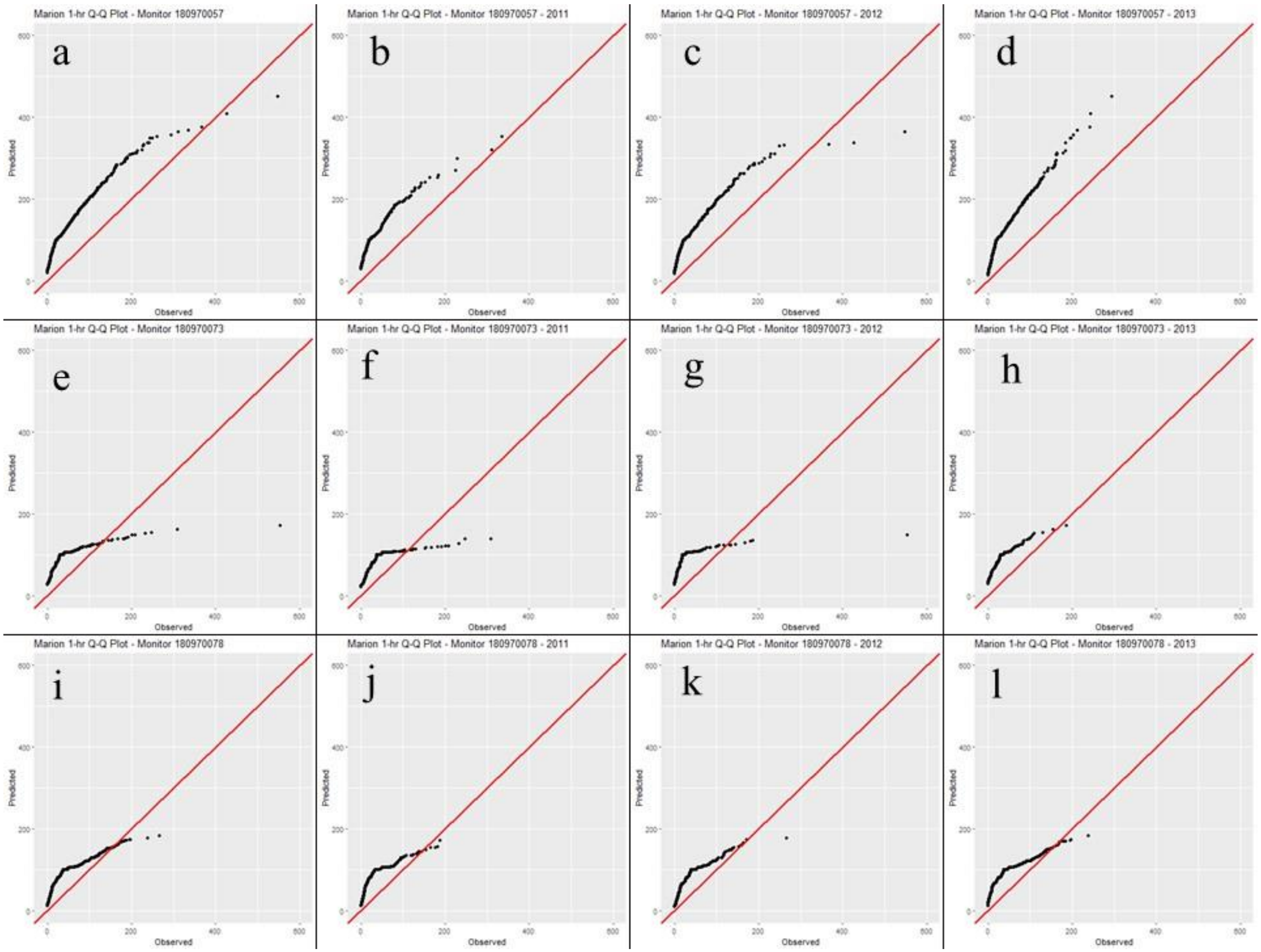


Figure D-7. 1-hour QQ plots for individual monitors in Indianapolis.

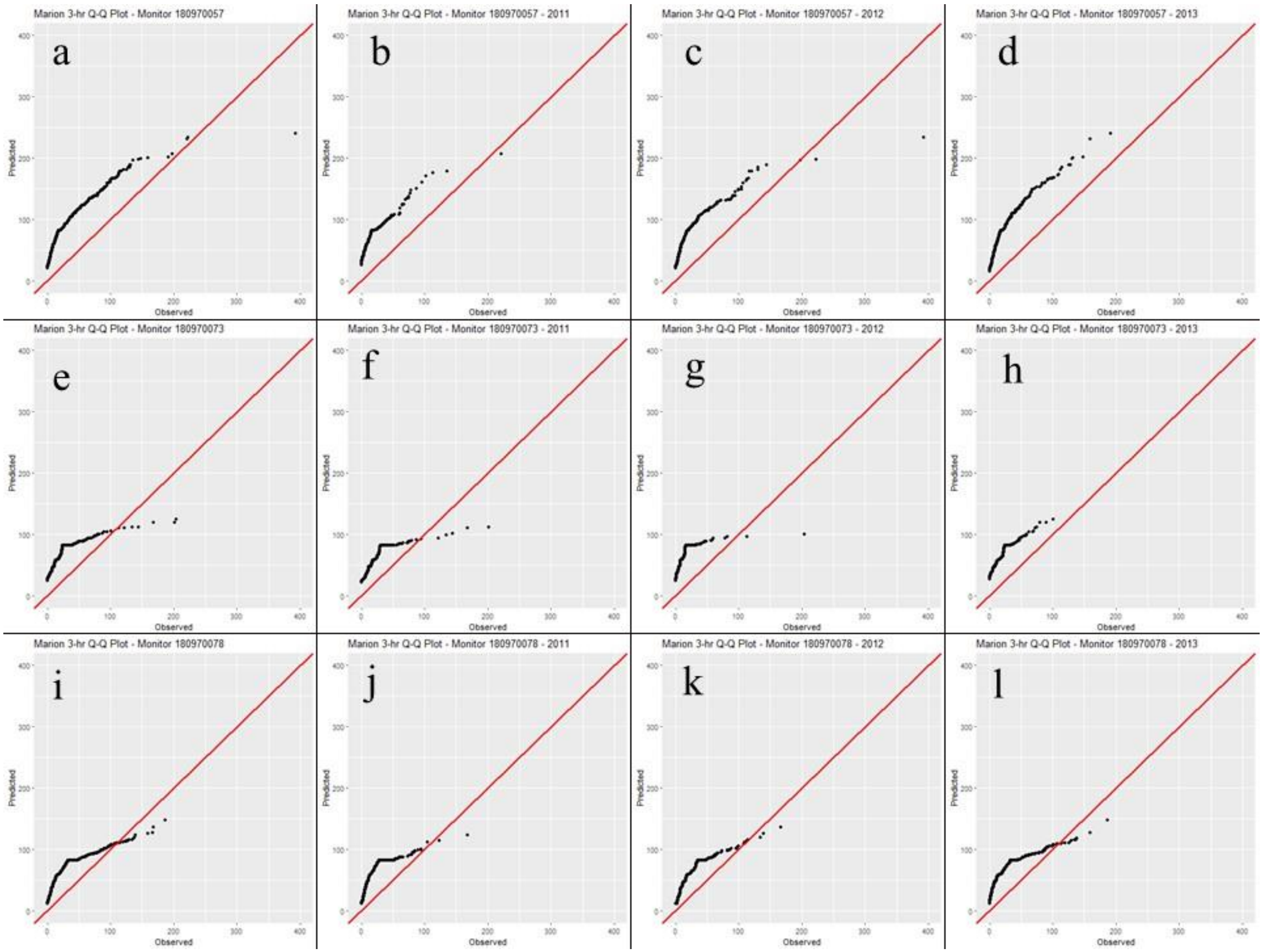


Figure D-8. 3-hour QQ-plots for individual monitors in Indianapolis.

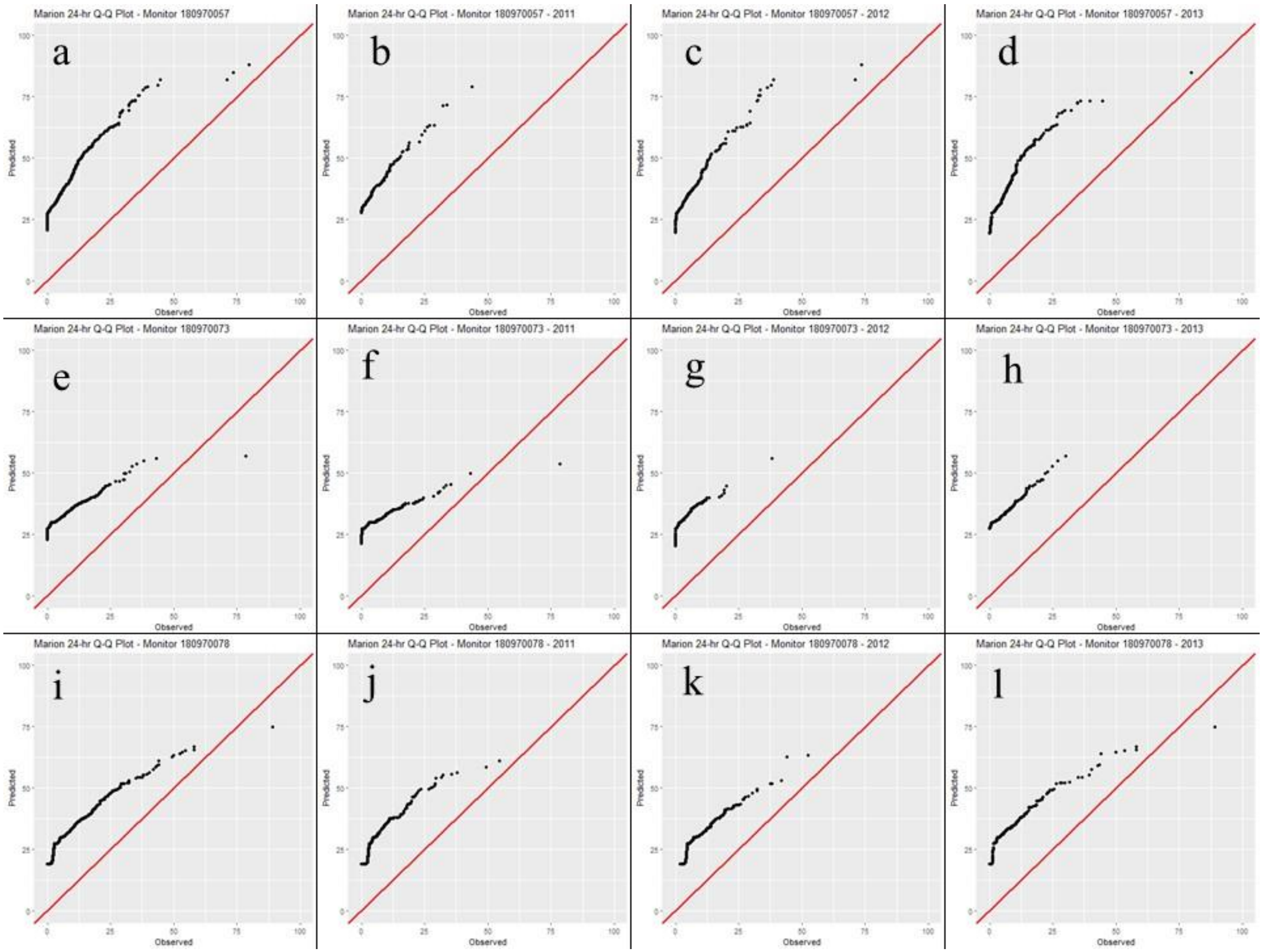


Figure D-9. 24-hour QQ-plots for individual monitors in Indianapolis.

CPM values were calculated for 2011, 2012, and 2013 and the entire 3-year period and are shown in Table D-4 across all monitors and each individual monitor.

Table D-4. Indianapolis composite performance metrics (CPM) and absolute fractional biases for 1-hour, 3-hour, and 24-hour averages.

Period	Monitor	CPM	$AFB_{1\text{-}hr}$	$AFB_{3\text{-}hr}$	$AFB_{24\text{-}hr}$
2011-2013	All	0.21	0.41	0.03	0.19
	180970057	0.33	0.62	0.03	0.34
	180970073	0.28	0.28	0.40	0.17
	180970078	0.25	0.35	0.37	0.03
2011	All	0.32	0.45	0.14	0.38
	180970057	0.48	0.61	0.26	0.57
	180970073	0.34	0.44	0.49	0.07
	180970078	0.26	0.29	0.20	0.30
2012	All	0.32	0.53	0.01	0.43
	180970057	0.37	0.66	0.01	0.43
	180970073	0.44	0.60	0.16	0.56
	180970078	0.23	0.34	0.21	0.15
2013	All	0.29	0.53	0.17	0.16
	180970057	0.45	0.74	0.17	0.43
	180970073	0.44	0.44	0.26	0.61
	180970078	0.24	0.41	0.29	0.03

The CPM values based on all monitors indicates relatively good model performance, for each individual year, as well as the entire 3-year period. Monitor 180970057 tends to have higher CPM values than the other monitors, possibly due to the inclusion of background increasing concentration while the monitor is impacted by most of the modeled sources as well. The one outlier in the CPM values is monitor 180970073 for 2011, with a CPM value of 0.74, much higher than the other monitors in 2011 or the CPM based on all three monitors. The high CPM appears to be due to the high AFB values for the 3-hour and 24-hour periods for the monitor as the monitor under-predicts compared to the other monitors for 2011 (Figures 3-11f and 3-12f).

Tulsa: Modeled Air Quality Evaluation

Three monitors were available for model evaluation in Tulsa (Figures C-5 and C-6). Table D-5 shows the annual 99th percentile of the daily 1-hour maximum concentrations and design values for each monitor. The model under-predicts the design value for 401430175 but does very well at the design value predictions for the other two monitors.

Table D-5. Tulsa monitored and modeled annual 99th percentile daily 1-hour maximum concentrations (μg m^{-3}) and 3-year design value (μg m^{-3}).

Monitor	Year	Observed	Modeled
401430175	2011	177.9	141.3
	2012	143.9	117.7
	2013	109.9	63.9
	Design Value	143.9	107.6
401430235	2011	88.9	122.8
	2012	62.8	99.7
	2013	49.8	52.6
	Design Value	67.1	91.7
401431127	2011	66.2	63.9
	2012	40.5	56.6
	2013	51.8	36.8
	Design Value	52.8	52.4

One-hour, 3-hour, and 24-hour average QQ-plots are shown in Figures D-10 through D-12 respectively across all monitors and QQ-plots by monitor are shown in Figures D-13 through D-15. For the 1-hour averages (Figure D-10), the model tends to over-predict for much of the concentration distribution for the total 3-year period as well as 2011 and 2012. 2013 shows a trend to more of the distribution being under-predicted. The 3-hour averages (Figure D-11) also show a trend of over-prediction and then under-prediction at the high end of the concentration distributions but perhaps less pronounced over-prediction than for the 1-hour averages. The 24-hour averages (Figure D-12) for the 3-year period show slight over-prediction at the lower ends of the distribution with good agreement in the middle followed by under-prediction but over-prediction at the very top of the distribution. 2011 shows slight over-prediction for much of the distribution, followed by under-prediction and over-prediction for the top three concentrations. 2012 and 2013 show mostly under-prediction, except at the lower end of the concentration distributions.

With regards to individual monitor performance, monitor 401430175 (located just north of the West Refinery in Figure C-5 and Figure C-6, appeared to have better model performance for the 1-hour averages based on the 1-hour QQ-plots (Figure D-13a) when considering the entire 3-year period. Monitor 401430175 under-predicted for 2011, a mix of under-prediction and slight over-prediction for 2012 and mostly over-prediction 2013. The other two monitors mostly over-predicted for the 3-year period and each individual year. For the 3-hour averages, monitor 4011431127 appeared to be the better performer (Figure D-14i-l) while monitor 401430175 tended toward over-prediction at the low end of the concentrations and under-prediction at the higher end. Monitor 401430235 mostly over-predicted. Similar trends for the monitors are seen in the 24-hour averages (Figure D-15).

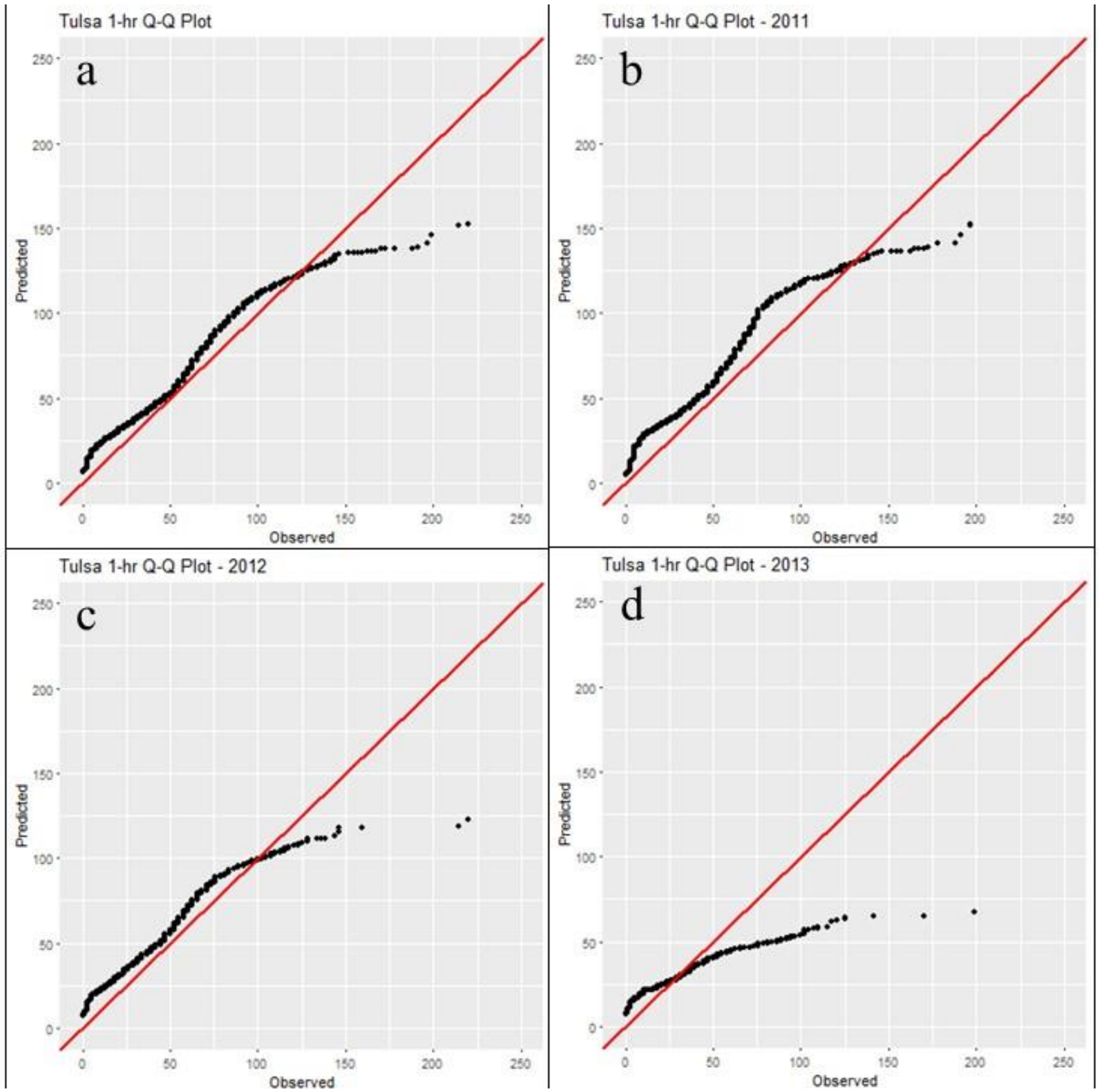


Figure D-10. Tulsa 1-hour QQ-plots.

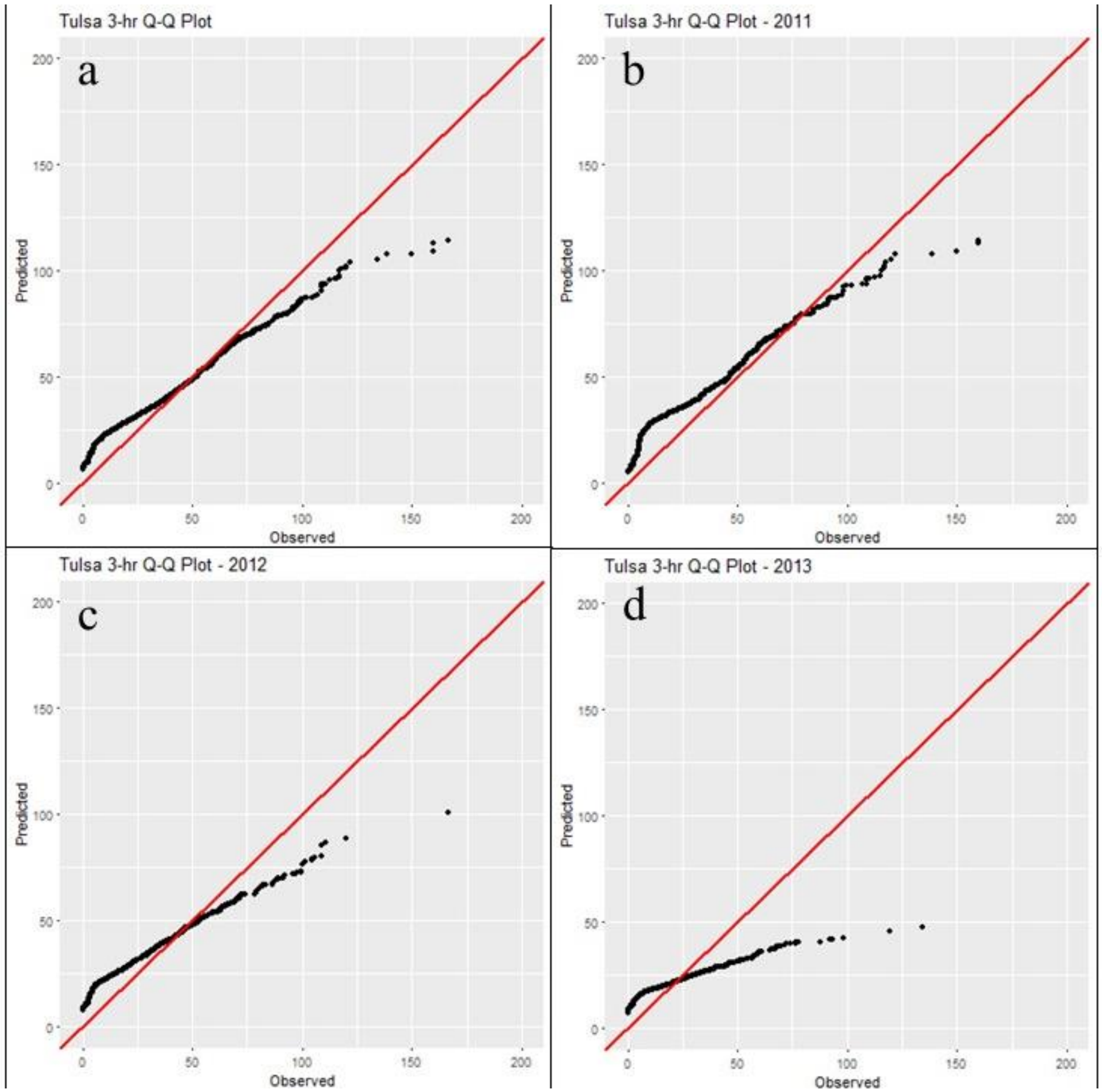


Figure D-11. Tulsa 3-hour QQ-plots.

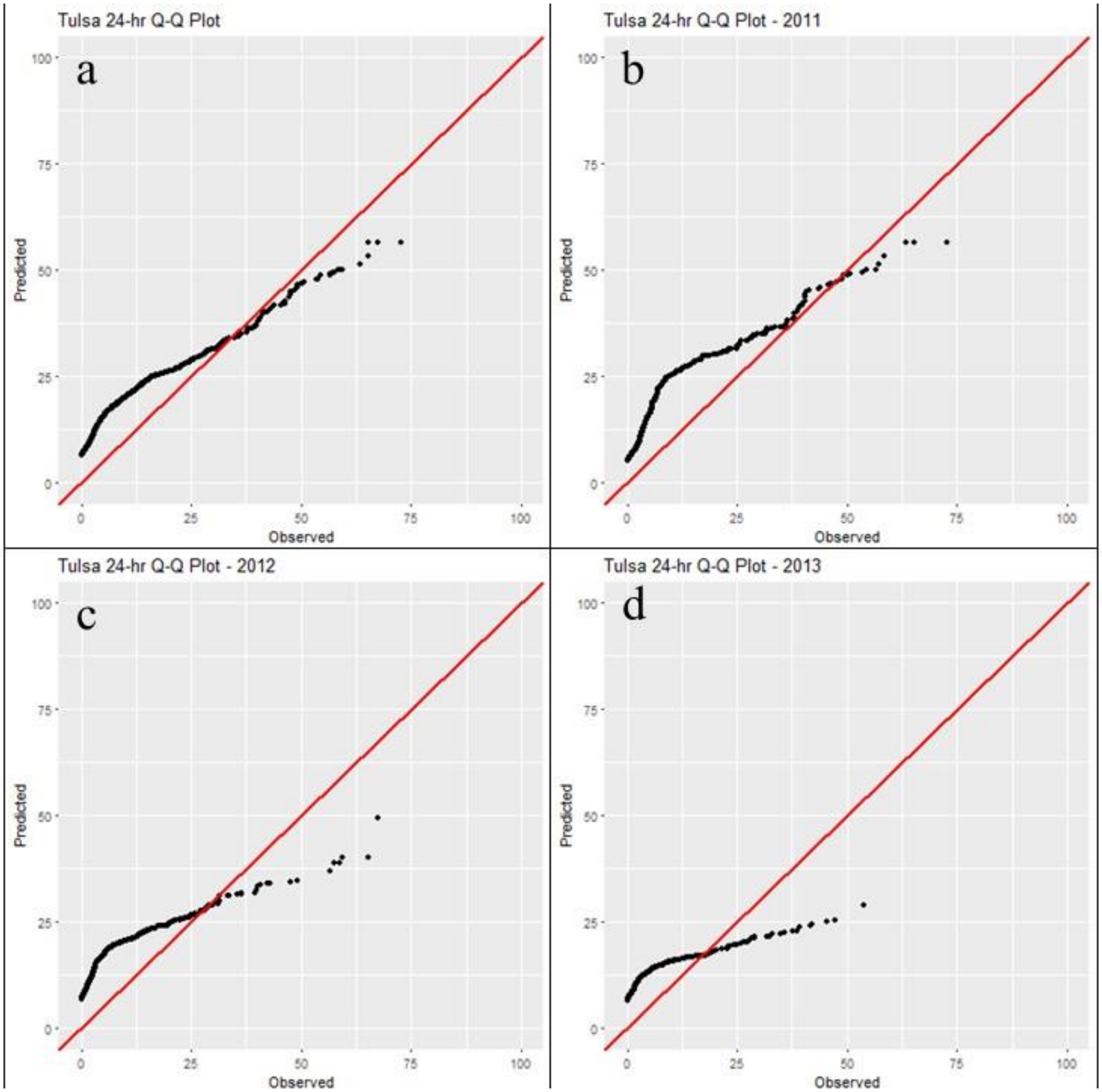


Figure D-12. Tulsa 24-hour QQ-plots.

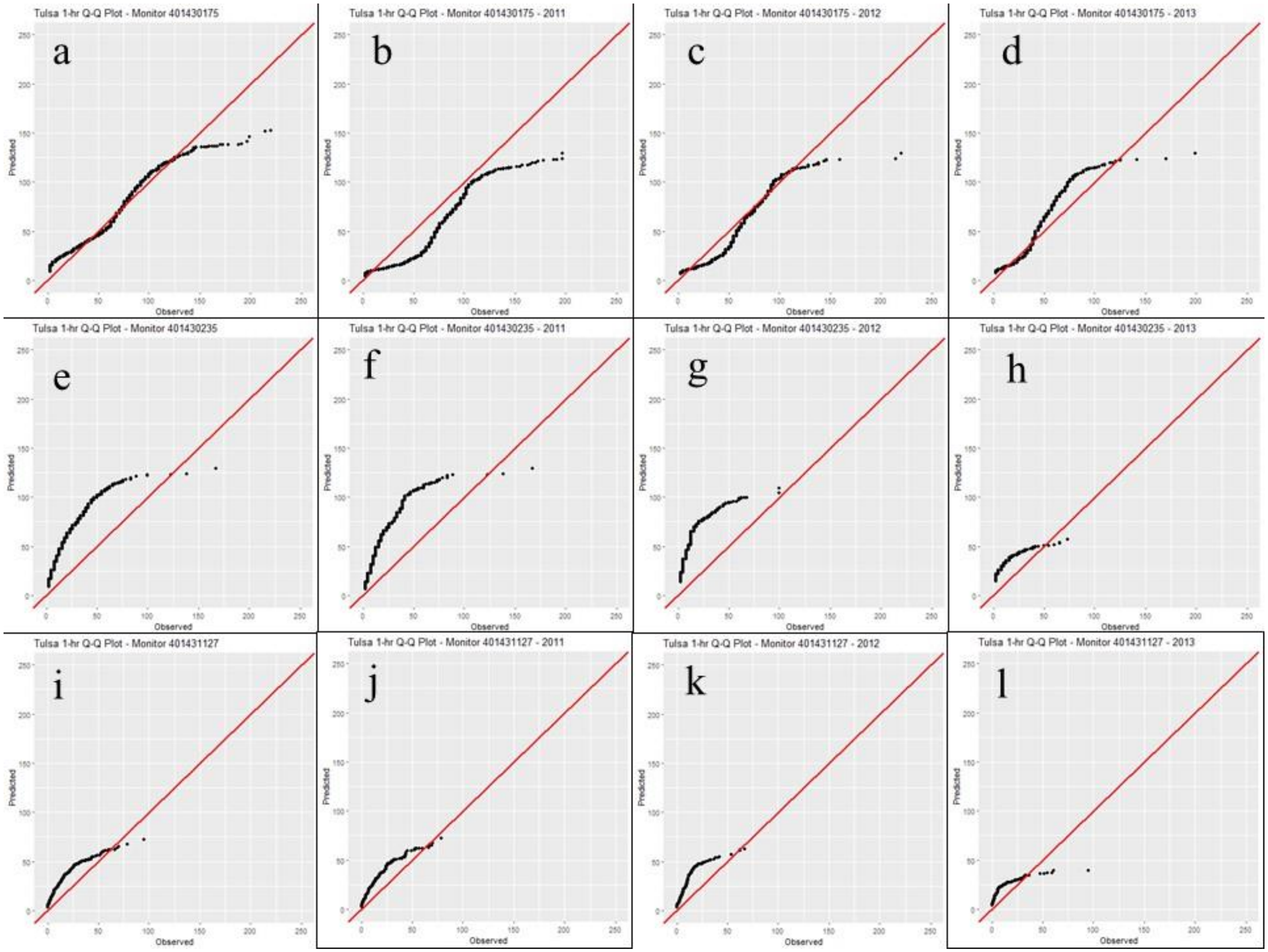


Figure D-13. 1-hour QQ-plots for individual monitors in Tulsa, OK.

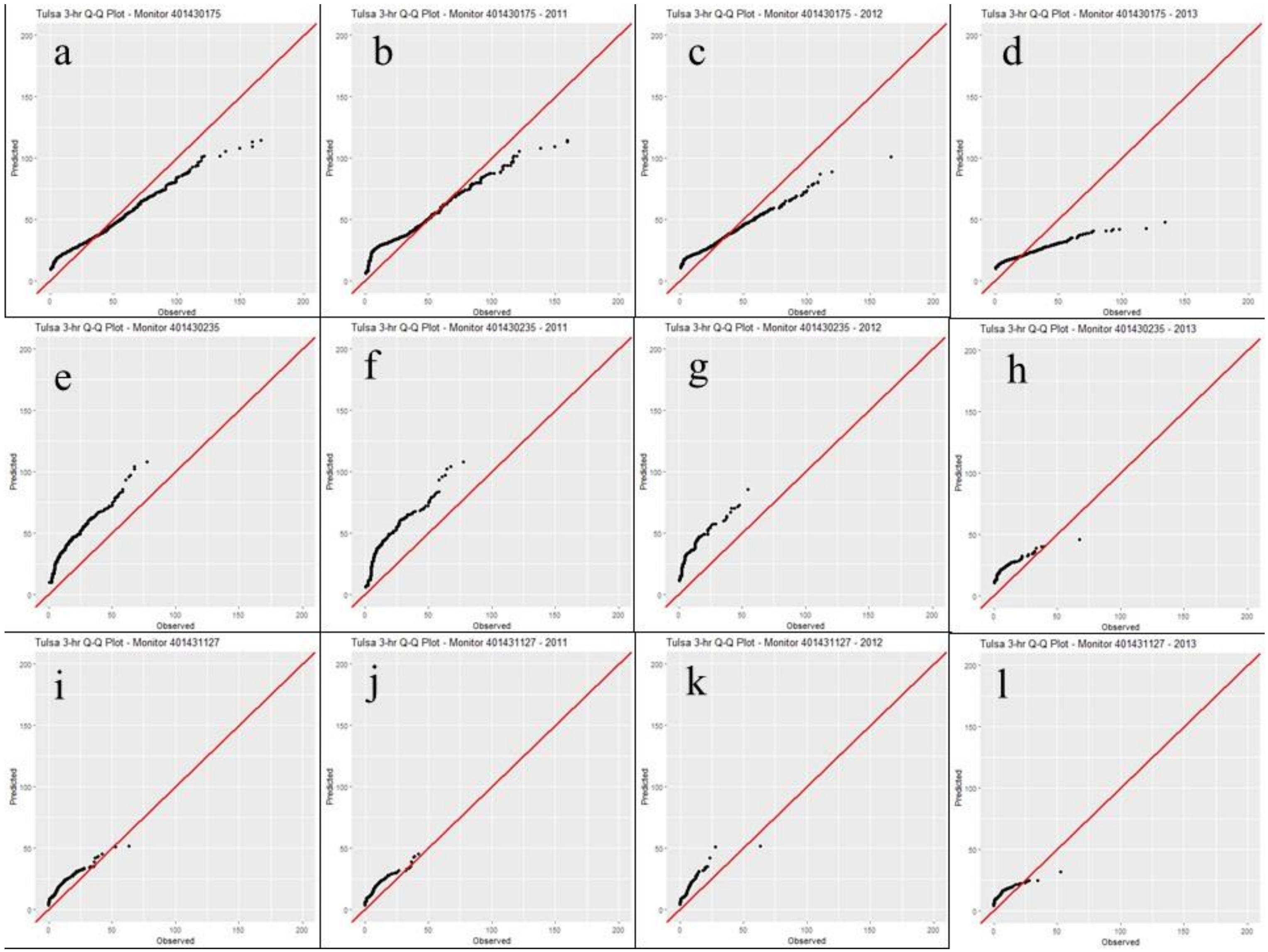


Figure D-14. 3-hour QQ-plots for individual monitors in Tulsa, OK.

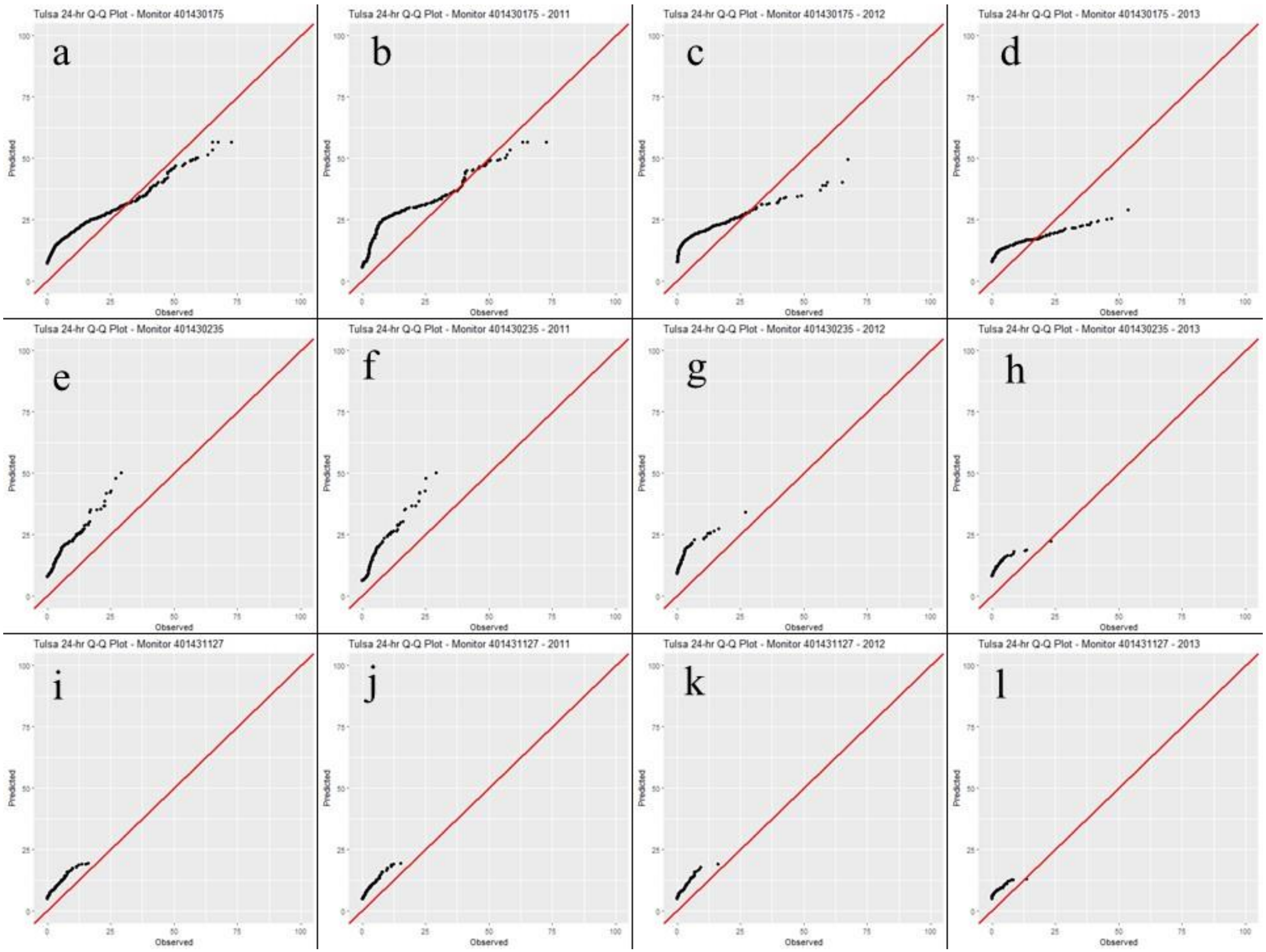


Figure D-15. 24-hour QQ-plots for individual monitors in Tulsa, OK.

CPM values were calculated for 2011, 2012, and 2013 and the entire 3-year period (Table D-6) across all monitors and each individual monitor. The CPM values among the individual monitors and the CPM based on all monitors tend to be very close to one another. The model with best agreement is 410431127 which tends to have the lower CPM with the exception of 4010432035 in 2013. Based on the CPM values, the model appears to do reasonably well against the monitored values, with the exception of 2013, where the high CPM of 401430175 is driving the overall CPM value across all monitors.

Table D-6. Tulsa composite performance metrics (CPM) and absolute fractional biases for 1-hour, 3-hour, and 24-hour averages.

Period	Monitor	CPM	AFB_{1-hr}	AFB_{3-hr}	AFB_{24-hr}
2011-2013	All	0.29	0.42	0.29	0.16
	401430175	0.34	0.57	0.29	0.16
	401432035	0.36	0.27	0.34	0.47
	410431127	0.31	0.42	0.18	0.33
2011	All	0.28	0.36	0.33	0.17
	401430175	0.34	0.52	0.33	0.17
	401432035	0.31	0.24	0.24	0.45
	410431127	0.29	0.32	0.14	0.41
2012	All	0.43	0.42	0.37	0.51
	401430175	0.49	0.59	0.37	0.51
	401432035	0.42	0.54	0.30	0.41
	410431127	0.34	0.13	0.34	0.55
2013	All	0.72	0.63	0.84	0.68
	401430175	0.83	0.97	0.84	0.68
	401432035	0.33	0.42	0.18	0.36
	410431127	0.37	0.50	0.37	0.24

Overall Model Performance Summary

Overall, for the three modeled areas, given uncertainties in emissions and meteorology and temporal resolution of the emissions for many of the sources (i.e., monthly, hour-of-day, month-hour-of-day, not individual hours), AERMOD appears to show adequate model performance, both from a regulatory evaluation standpoint, and the narrower analysis on a monitor-by-monitor-basis. When evaluating on an annual basis, 2011 tended to be the better performing year, which is not surprising given that 2011 is one of the triennial emissions inventory years. Also, as noted, given the temporal resolution of the most of the emissions, the model performance is quite good. With some of the sources using a monthly temporal profile, emissions for each hour for a given month would be the same (See Appendix B of this document for an example). Given the lack of temporal variability of source emissions in the model and the fact that a monitor does pick up temporal variability of emissions not seen by the model, the performance of AERMOD is acceptable for the purposes of this exposure assessment.

REFERENCES

Cox, W.M. and J.A. Tikvart. (1990). A statistical Procedure for Determining the Best Performing Air Quality Simulation Model. *Atmos. Environ.*, **24A (9)**: 2387-2395.

U.S. EPA. (1992). Protocol for Determining the Best Performing Model, EPA-454/R-92-025. U.S. Environmental Protection Agency, Research Triangle Park, NC.

U.S. EPA. (2003). AERMOD: Latest Features and Evaluation Results. EPA-454/R-03-003. U.S. Environmental Protection Agency, Research Triangle Park, NC 27711.

U.S. EPA. (2016). Evaluation of Prognostic Meteorological Data in AERMOD Applications, EPA-454/R-16-004. U.S. Environmental Protection Agency, Research Triangle Park, North Carolina 27711.

Venkatram, A., R. W. Brode, A. J. Cimorelli, J. T. Lee, R. J. Paine, S. G. Perry, W. D. Peters, J. C. Weil, and R. B. Wilson. (2001). A complex terrain dispersion model for regulatory applications. *Atmos.Environ.*, 35, 4211-4221.

APPENDIX E

ASTHMA PREVALENCE

E.1 Overview

This appendix describes the development of the most recent asthma prevalence file used by EPA's Air Pollution Exposure Model (APEX) to estimate individuals (e.g., children, adults) having asthma. This development involved three basic steps: 1) processing National Health Interview Survey (NHIS) asthma prevalence data, 2) processing U.S. Census poverty/income status data, and 3) combining the two sets considering variables known to influence asthma (e.g., age, sex, poverty status, U.S. region) to estimate asthma prevalence stratified by age and sex for all US Census tracts.

E.2 General History

The current processing approach is based on work originally performed by Cohen and Rosenbaum (2005) and then revised and extended by U.S. EPA (2014). Briefly for the earlier APEX asthma prevalence file development, Cohen and Rosenbaum (2005) calculated asthma prevalence for children aged 0 to 17 years for each age, sex, and four U.S. regions using 2003 NHIS survey data. The regions defined by NHIS were 'Midwest', 'Northeast', 'South', and 'West'. The asthma prevalence was defined as the probability of a 'Yes' response to the question "EVER been told that [the child] had asthma?"[1] among those persons that responded either 'Yes' or 'No' to this question.[2] The responses were weighted to take into account the complex survey design of the NHIS.[3] Standard errors and confidence intervals for the prevalence were calculated using a logistic model (PROC SURVEY LOGISTIC). A scatterplot technique (LOESS smoother) was applied to smooth the prevalence curves and compute the standard errors and confidence intervals for the smoothed prevalence estimates. Logistic analysis of the raw and smoothed prevalence curves showed statistically significant differences in prevalence by gender and region, supporting their use as stratification variables in the final data set. These smoothed prevalence estimates were used as an input to APEX to estimate air pollutant exposure in children with asthma (U.S. EPA 2007; 2008; 2009).

---

[1] The response was recorded as variable "CASHMEV" in the downloaded dataset. Data and documentation are available at http://www.cdc.gov/nchs/nhis/quest_data_related_1997_forward.htm.

[2] If there were another response to this variable other than "yes" or "no" (i.e., refused, not ascertained, don't know, and missing), the surveyed individual was excluded from the analysis data set.

[3] In the SURVEY LOGISTIC procedure, the variable "WTF_SC" was used for weighting, "PSU" was used for clustering, and "STRATUM" was used to define the stratum.

In the revision documented in U.S. EPA (2014), several years of NHIS survey data (2006-2010) were combined and used to calculate asthma prevalence for that period. Asthma prevalence for children (by age in years) as was estimated as described above but also included an estimate of adult asthma prevalence (by age groups). In addition, two sets of asthma prevalence for each adults and children were estimated. The first data set, as was done previously, was based on responses to the question "EVER been told that [the child] had asthma". The second data set was developed using the probability of a 'Yes' response to a question that followed those that answered 'Yes' to the first question regarding ever having asthma, specifically, do those persons "STILL have asthma?". And finally, in addition to the nominal variables region and sex, the asthma prevalence in this new analysis were further stratified by a family income/poverty ratio (i.e., whether the family income was considered below or at/above the US Census estimate of poverty level for the given year).

These updated asthma prevalence data were linked to U.S. census tract level poverty ratios probabilities, also stratified by age. Staff considered the variability in population exposures to be better represented when accounting for and modeling these newly refined attributes of this susceptible population. This is because of the 1) significant observed differences in asthma prevalence by age, sex, region, and poverty status, 2) the variability in the spatial distribution of poverty status across census tracts, stratified by age, and 3) the potential for spatial variability in local scale ambient concentrations.

It is in this spirit that staff update the asthma prevalence files used by APEX, using the most recent data available that reasonably bound the exposure assessment period of interest.

Step 1: NHIS Data Set Description and Processing

The objective of this first processing step was to estimate asthma prevalence for children and adults considering several influential variables. First, raw 2011-2015 data and associated documentation were downloaded from the Center for Disease Control (CDC) and Prevention's NHIS website.[4] The 'Sample Child' and 'Sample Adult' files were selected because of the availability of person-level attributes of interest within these files, i.e., age in years ('age_p'), sex ('sex'), U.S. geographic region ('region'), coupled with the response to questions of whether or not the surveyed individual ever had and still has asthma. In total, five years of recent survey data were obtained, comprising over 64,000 children and 170,000 children for years 2011-2015 (Table E-1).

[4] See http://www.cdc.gov/nchs/nhis.htm (accessed April 11, 2017).

Information regarding personal and family income and poverty ranking are also provided by the NHIS in separate files. Five files (‘INCIMPx.dat’) are available for each survey year, each containing either the actual responses (where recorded or provided by survey participant) or imputed values for the desired financial variable.[5] For this current analysis, the ratio of income to poverty was provided as a continuous variable (‘POVRATI3’) and used to develop a nominal variable for this evaluation: either the survey participant was below or above a selected poverty threshold. This was done in this manner to be consistent with data generated as part of the second data set processing step, i.e., a table containing census tract level poverty ratio probabilities stratified by age (step 2).

When considering the number of stratification variables, the level of asthma prevalence, and poverty distribution among the survey population, sample size was an important issue. For the adult data, there were insufficient numbers of persons available to stratify the data by single ages (for some years of age there were no survey persons). Therefore, the adult survey data were grouped as follows: ages 18-24, 25-34, 35-44, 45-54, 55-64, 65-74, and, ≥75.[6] To increase the number of persons within the age, gender, and four region groupings of our characterization of ‘below poverty’ asthmatics persons, the poverty ratio threshold was selected as <1.5, therefore including persons that were within 50% above the poverty threshold. If the mean of the five imputed/recorded values were <1.5, the person’s family income was categorized ‘below’ the poverty threshold, if the mean of the 5 values were ≥1.5, the person’s family income was categorized ‘above’ the poverty threshold.

The person-level income files were then merged with the sample adult and child files using the ‘HHX’ (a household identifier), ‘FMX’ (a family identifier), and ‘FPX’ (an individual identifier) variables. Note, all persons within the sample adult and child files had corresponding financial survey data.

Two asthma survey response variables were of interest in this analysis and were used to develop the two separate prevalence data sets for each children and adults. The response to the first question “Have you EVER been told by a doctor or other health professional that you [or your child] had asthma?” was recorded as variable name ‘CASHMEV’ for children and ‘AASMEV’ for adults. Only persons having responses of either ‘Yes’ or ‘No’ to this question were retained to estimate the asthma prevalence. This assumes that the exclusion of those

[5] Financial information was not collected from all persons; therefore, the NHIS provides imputed data. Details into the available variables and imputation method are provided with each year’s data set. For example, see “Multiple Imputation of Family Income and Personal Earnings in the National Health Interview Survey: Methods and Examples” at https://www.cdc.gov/nchs/data/nhis/tecdoc15.pdf.

[6] These same age groupings were used to create the companion file containing the census tract level poverty ratio probabilities (section 2).

responding otherwise, i.e., those that ‘refused’ to answer, instances where it was “not ascertained’, or the person ‘does not know’, does not affect the estimated prevalence rate if either ‘Yes’ or ‘No’ answers could actually be given by these persons. There were very few persons providing an unusable response (Table E-1), thus the above assumption is reasonable. A second question was asked as a follow to persons responding “Yes” to the first question, specifically, “Do you STILL have asthma?” and noted as variables ‘CASSTILL’ and ‘AASSTILL’ for children and adults, respectively. Again, while only persons responding ‘Yes’ and ‘No’ were retained for further analysis, the representativeness of the screened data set is assumed unchanged from the raw survey data given the few persons having unusable data.

Table E-1. Number of total surveyed persons from NHIS (2006-2010) sample adult and child files and the number of those responding to asthma survey questions.

CHILDREN	2011	2012	2013	2014	2015	TOTAL
All Persons	12,844	13,275	12,860	13,380	12,281	64,640
Yes/No Asthma	12,831	13,263	12,851	13,366	12,269	64,580
Yes/No to Still Have + No Asthma	12,831	13,248	12,844	13,359	12,269	64,551
ADULTS	**2011**	**2012**	**2013**	**2014**	**2015**	**TOTAL**
All Persons	33,014	34,525	34,557	36,697	33,672	172,465
Yes/No Asthma	32,982	34,505	34,525	36,667	33,651	172,330
Yes/No to Still Have + No Asthma	32,953	34,468	34,498	36,615	33,614	172,148

Logistic Models

As described in the previous section, four person-level analytical data sets were created from the raw NHIS data files, generally containing similar variables: a ‘Yes’ or ‘No’ asthma response variable (either ‘EVER’ or ‘STILL’), an age (or age group for adults), their sex (‘male’ or ‘female’), US geographic region (‘Midwest’, ‘Northeast’, ‘South’, and ‘West’), and poverty status (‘below’ or above’). One approach to calculate prevalence rates and their uncertainties for a given gender, region, poverty status, and age is to calculate the proportion of ‘Yes’ responses among the ‘Yes’ and ‘No’ responses for that demographic group, appropriately weighting each response by the survey weight. This simplified approach was initially used to develop ‘raw’ asthma prevalence rates however this approach may not be completely appropriate. The two main issues with such a simplified approach are that the distributions of the estimated prevalence rates would not be well approximated by normal distributions and that the estimated confidence intervals based on a normal approximation would often extend outside the [0, 1] interval. A better approach for such survey data is to use a logistic transformation and fit the model:

Prob (asthma) = exp(beta) / (1 + exp(beta)), where beta may depend on the explanatory variables for age, sex, poverty status, or region. This is equivalent to the model:

$$\text{Beta} = \text{logit}\ \{\text{prob (asthma)}\} = \log\ \{\text{prob (asthma)} / [1 - \text{prob (asthma)}]\}.$$

The distribution of the estimated values of beta is more closely approximated by a normal distribution than the distribution of the corresponding estimates of prob (asthma). By applying a logit transformation to the confidence intervals for beta, the corresponding confidence intervals for prob (asthma) will always be inside [0, 1]. Another advantage of the logistic modeling is that it can be used to compare alternative statistical models, such as models where the prevalence probability depends upon age, region, poverty status, and sex, or on age, region, poverty status but not sex.

In previous analyses using the 2006-2010 NHIS asthma prevalence data, a variety of logistic models and compared them for use in estimating asthma prevalence, where the transformed probability variable beta is a given function of age, gender, poverty status, and region (Cohen and Rosenbaum, 2005; U.S. EPA, 2014). The SAS procedure SURVEYLOGISTIC was used to fit the various logistic models, taking into account the NHIS survey weights and survey design (using both stratification and clustering options), as well as considering various combinations of the selected explanatory variables.

As an example, Table E-2 lists the models fit and their log-likelihood goodness-of-fit measures using the sample child data and for the “EVER” asthma response variable using the 2006-2010 NHIS data. A total of 32 models were fit, depending on the inclusion of selected explanatory variables and how age was considered in the model. The ‘Strata’ column lists the eight possible stratifications: no stratification, stratified by gender, by region, by poverty status, by region and gender, by region and poverty status, by gender and poverty status, and by region, gender and poverty status. For example, “5. region, gender” indicates that separate prevalence estimates were made for each combination of region and gender. As another example, “2. gender” means that separate prevalence estimates were made for each gender, so that for each gender, the prevalence is assumed to be the same for each region. Note the prevalence estimates are independently calculated for each stratum.

The ‘Description’ column of Table E-2 indicates how beta depends upon the age:

Linear in age	Beta = $\alpha + \beta \times$ age, where α and β vary with strata.
Quadratic in age	Beta = $\alpha + \beta \times \text{age} + \gamma \times \text{age}^2$ where α β and γ vary with strata.

Cubic in age	Beta = $\alpha + \beta \times age + \gamma \times age^2 + \delta \times age^3$ where α β, γ, and δ vary with the strata.
f(age)	Beta = arbitrary function of age, with different functions for different strata

The category *f(age)* is equivalent to making age one of the stratification variables, and is also equivalent to making beta a polynomial of degree 17 in age (since the maximum age for children is 17), with coefficients that may vary with the strata.

The fitted models are listed in order of complexity, where the simplest model (1) is a non-stratified linear model in age and the most complex model (model 32) has a prevalence that is an arbitrary function of age, gender, poverty status, and region. Model 32 is equivalent to calculating independent prevalence estimates for each of the 288 combinations of age, sex, poverty status, and region.

Table E-2. Alternative logistic models for estimating child asthma prevalence using the "EVER" asthma response variable and goodness of fit test results using the 2006-2010 NHIS data.

Model	Description	Strata	- 2 Log Likelihood	DF
1	1. logit(prob) = linear in age	1. none	288740115.1	2
2	1. logit(prob) = linear in age	2. gender	287062346.4	4
3	1. logit(prob) = linear in age	3. region	288120804.1	8
4	1. logit(prob) = linear in age	4. poverty	287385013.1	4
5	1. logit(prob) = linear in age	5. region, gender	286367652.6	16
6	1. logit(prob) = linear in age	6. region, poverty	286283543.6	16
7	1. logit(prob) = linear in age	7. gender, poverty	285696164.7	8
8	1. logit(prob) = linear in age	8. region, gender, poverty	284477928.1	32
9	2. logit(prob) = quadratic in age	1. none	286862135.1	3
10	2. logit(prob) = quadratic in age	2. gender	285098650.6	6
11	2. logit(prob) = quadratic in age	3. region	286207721.5	12
12	2. logit(prob) = quadratic in age	4. poverty	285352164	6
13	2. logit(prob) = quadratic in age	5. region, gender	284330346.1	24
14	2. logit(prob) = quadratic in age	6. region, poverty	284182547.5	24
15	2. logit(prob) = quadratic in age	7. gender, poverty	283587631.7	12
16	2. logit(prob) = quadratic in age	8. region, gender, poverty	282241318.6	48
17	3. logit(prob) = cubic in age	1. none	286227019.6	4
18	3. logit(prob) = cubic in age	2. gender	284470413	8
19	3. logit(prob) = cubic in age	3. region	285546716.1	16
20	3. logit(prob) = cubic in age	4. poverty	284688169.9	8
21	3. logit(prob) = cubic in age	5. region, gender	283662673.5	32
22	3. logit(prob) = cubic in age	6. region, poverty	283404487.5	32
23	3. logit(prob) = cubic in age	7. gender, poverty	282890785.3	16
24	3. logit(prob) = cubic in age	8. region, gender, poverty	281407414.3	64
25	4. logit(prob) = f(age)	1. none	285821686.2	18
26	4. logit(prob) = f(age)	2. gender	283843266.2	36
27	4. logit(prob) = f(age)	3. region	284761522.8	72
28	4. logit(prob) = f(age)	4. poverty	284045849.2	36
29	4. logit(prob) = f(age)	5. region, gender	282099156.1	144
30	4. logit(prob) = f(age)	6. region, poverty	281929968.5	144
31	4. logit(prob) = f(age)	7. gender, poverty	281963915.7	72
32	4. logit(prob) = f(age)	8. region, gender, poverty	278655423.1	288

Table E-2 also includes the -2 Log Likelihood statistic, a goodness-of-fit measure, and the associated degrees of freedom (DF), which is the total number of estimated parameters. Any two models can be compared using their -2 Log Likelihood values: models having lower values are preferred. If the first model is a special case of the second model, then the approximate statistical significance of the first model is estimated by comparing the difference in the -2 Log Likelihood values with a chi-squared random variable having *r* degrees of freedom, where *r* is the difference in the DF (hence a likelihood ratio test). For all pairs of models from Table E-2, all the differences in the -2 Log Likelihood statistic are at least 600,000 and thus significant at p-values well below 1 percent. Based on its having the lowest -2 Log Likelihood value, the last model fit (model 32: retaining all explanatory variables and using *f(age)*) was preferred and used to estimate the asthma prevalence in the prior analyses[7] as well as employed for this updated 2011-2015 NHIS data analysis.

The SURVEYLOGISTIC procedure produces estimates of the beta values and their 95% confidence intervals for each combination of age, region, poverty status, and gender. By applying the inverse logit transformation,

$$\text{Prob (asthma)} = \exp(\text{beta}) / (1 + \exp(\text{beta})),$$

one can convert the beta values and associated 95% confidence intervals into predictions and 95% confidence intervals for the prevalence. The standard error for the prevalence was estimated as:

$$\text{Std Error \{Prob (asthma)\}} = \text{Std Error (beta)} \times \exp(-\text{beta}) / (1 + \exp(\text{beta}))^2,$$

which follows from the delta method (i.e., a first order Taylor series approximation).

Estimated asthma prevalence using this approach and termed here as ‘unsmoothed’ are provided in Attachment 1. Graphical representation is provided in a series of figures incorporating the following variables:

- Region
- Gender
- Age (in years) or Age_group (age categories)

[7] Similar results were obtained when estimating prevalence using the ‘STILL’ have asthma variable as well as when investigating model fit using the adult data sets. In the Cohen and Rosenbaum (2005) analysis, adult data were not used and the poverty to income ratio was not a variable in their models. Also, because age was a categorical variable in the adult data sets in U.S. EPA (2014) and analyses conducted here, it could only be evaluated using f(age_group).

- Poverty Status
- Prevalence = predicted prevalence
- SE = standard error of predicted prevalence
- LowerCI = lower bound of 95 % confidence interval for predicted prevalence
- UpperCI = upper bound of 95 % confidence interval for predicted prevalence

A series of 8 plots are provided per figure that vary by region and poverty status (i.e., 4 x 2 = 8). Results for children are given in Figures 1 ('EVER' had Asthma) and 2 ('STILL' have asthma) while adults are provided in Figures 3 ('EVER' had Asthma) and 4 ('STILL' have asthma) within Attachment 1. Data used for each figure/plot can be provided upon request.

Loess Smoother

The estimated prevalence curves show that the prevalence is not necessarily a smooth function of age. The linear, quadratic, and cubic functions of age modeled by SURVEYLOGISTIC were identified as a potential method for smoothing the curves, but they did not provide the best fit to the data. One reason for this might be due to the attempt to fit a global regression curve to all the age groups, which means that the predictions for age *A* are affected by data for very different ages. A local regression approach that separately fits a regression curve to each age A and its neighboring ages was used, giving a regression weight of 1 to the age *A*, and lower weights to the neighboring ages using a tri-weight function:

$$\text{Weight} = \{1 - [\,|\text{age} - A| / q\,]^3\}, \text{ where } |\text{age} - A| <= q.$$

The parameter *q* defines the number of points in the neighborhood of the age *A*. Instead of calling *q* the smoothing parameter, SAS defines the smoothing parameter as the proportion of points in each neighborhood. A quadratic function of age to each age neighborhood was fit separately for each gender and region combination. These local regression curves were fit to the beta values, the logits of the asthma prevalence estimates, and then converted them back to estimated prevalence rates by applying the inverse logit function exp(beta) / (1 + exp(beta)). In addition to the tri-weight variable, each beta value was assigned a weight of
$1 / [\text{std error (beta)}]^2$, to account for their uncertainties.

In this application of LOESS, weights of $1 / [\text{std error (beta)}]^2$ were used such that $\sigma^2 = 1$. The LOESS procedure estimates σ^2 from the weighted sum of squares. Because it is assumed $\sigma^2 = 1$, the estimated standard errors are multiplied by 1 / estimated σ and adjusted the widths of the confidence intervals by the same factor.

There are several potential values that can be selected for the smoothing parameter; the optimum value was determined by evaluating three regression diagnostics: the residual standard error, normal probability plots, and studentized residuals. To generate these statistics, the LOESS procedure was applied to estimated smoothed curves for beta, the logit of the prevalence, as a function of age, separately for each region, gender, and poverty classification. For the children data sets, curves were fit using the choices of 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, and 1.0 for the smoothing parameter. This selected range of values was bounded using the following observations. With only 18 points (i.e., the number of single year ages for children), a smoothing parameter of 0.2 cannot be used because the weight function assigns zero weights to all ages except age *A*, and a quadratic model cannot be uniquely fit to a single value. A smoothing parameter of 0.3 also cannot be used because that choice assigns a neighborhood of 5 points only ($0.3 \times 18 = 5$, rounded down), of which the two outside ages have assigned weight zero, making the local quadratic model fit exactly at every point except for the end points (ages 0, 1, 16 and 17). Usually one uses a smoothing parameter below 1 so that not all the data are used for the local regression at a given x value. Note also that a smoothing parameter of 0 can be used to generate the raw, unsmoothed, prevalence. The selection of the smoothing parameter used for the adult curves would follow a similar logic, although the lower bound could effectively be extended only to 0.9 given the number of age groups. This limits the selection of smoothing parameter applied to the two adult data sets to a value of 0.9, though values of 0.8 – 1.0 were nevertheless compared for good measure.

The first regression diagnostic used was the residual standard error, which is the LOESS estimate of σ. As discussed above, the true value of σ equals 1, so the best choice of smoothing parameter should have residual standard errors as close to 1 as possible. For children 'EVER' having asthma and when considering the best models (of the 112 possible, those having $0.95<RSE<1.05$) using this criterion, the best choice varies with gender, region, and poverty status between smoothing parameters of 0.4, 0.7, and 1.0 (Table E-3). For the 'STILL' data set, a value of 0.5 or 0.6 would be slightly preferred. The 'EVER' adult data set could be smoothed using a value of 0.8 – 1.0 given the limited selection of smoothing values (of the 48 possible models), though 0.8 appears a better value for the 'STILL' data set.

Table E-3. Top model smoothing fits where residual standard error at or a value of 1.0.

Data Set	Asthma	Smoothing Parameter						
		0.4	0.5	0.6	0.7	0.8	0.9	1.0
Children	EVER	4	2	2	4	3	3	4
	STILL	3	5	4	2	3	2	2
Adults	EVER	n/a	n/a	n/a	n/a	3	3	5
	STILL	n/a	n/a	n/a	n/a	3	1	1

The second regression diagnostic was developed from an approximate studentized residual. The residual errors from the LOESS model were divided by standard error (beta) to make their variances approximately constant. These approximately studentized residuals should be approximately normally distributed with a mean of zero and a variance of $\sigma^2 = 1$. To test this assumption, normal probability plots of the residuals were created for each smoothing parameter, combining all the studentized residuals across genders, regions, poverty status, and ages. The results for the children data indicate little distinction or affect by the selection of a particular smoothing parameter (e.g., see Figure E-1), although linearity in the plotted curve is best expressed with smoothing parameters generally between 0.6 and 0.9. When considering the adult data sets, the appropriate value would generally be 0.9.

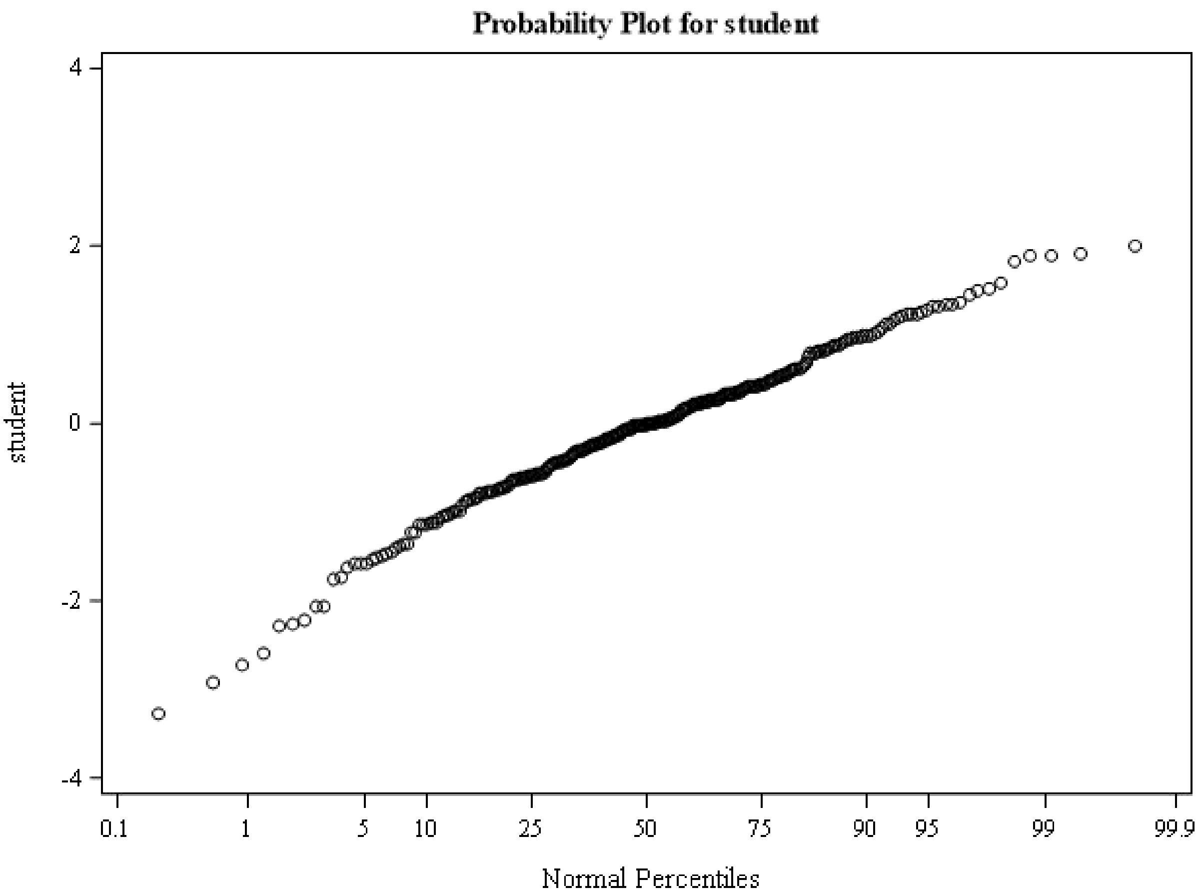


Figure E-1. Normal probability plot of studentized residuals generated using logistic model, smoothing set to 0.6, and the children 'STILL' asthmatic data set.

The third regression diagnostic are plots of the studentized residuals against the smoothed beta values. All the studentized residuals for a given smoothing parameter are plotted together within the same graph. Also plotted is a LOESS smoothed curve fit to the same set of points, with SAS's optimal smoothing parameter choice, to indicate the typical pattern. Ideally there should be no obvious pattern and an average studentized residual close to zero with no regression slope (e.g., see Figure E-2). For the children data sets, these plots generally indicate no unusual patterns, and the results for smoothing parameters 0.4 through 0.6 indicate a fit LOESS curve closest to the studentized residual equals zero line. When considering the adult data sets, 0.9 – 1.0 appear to be appropriate values.

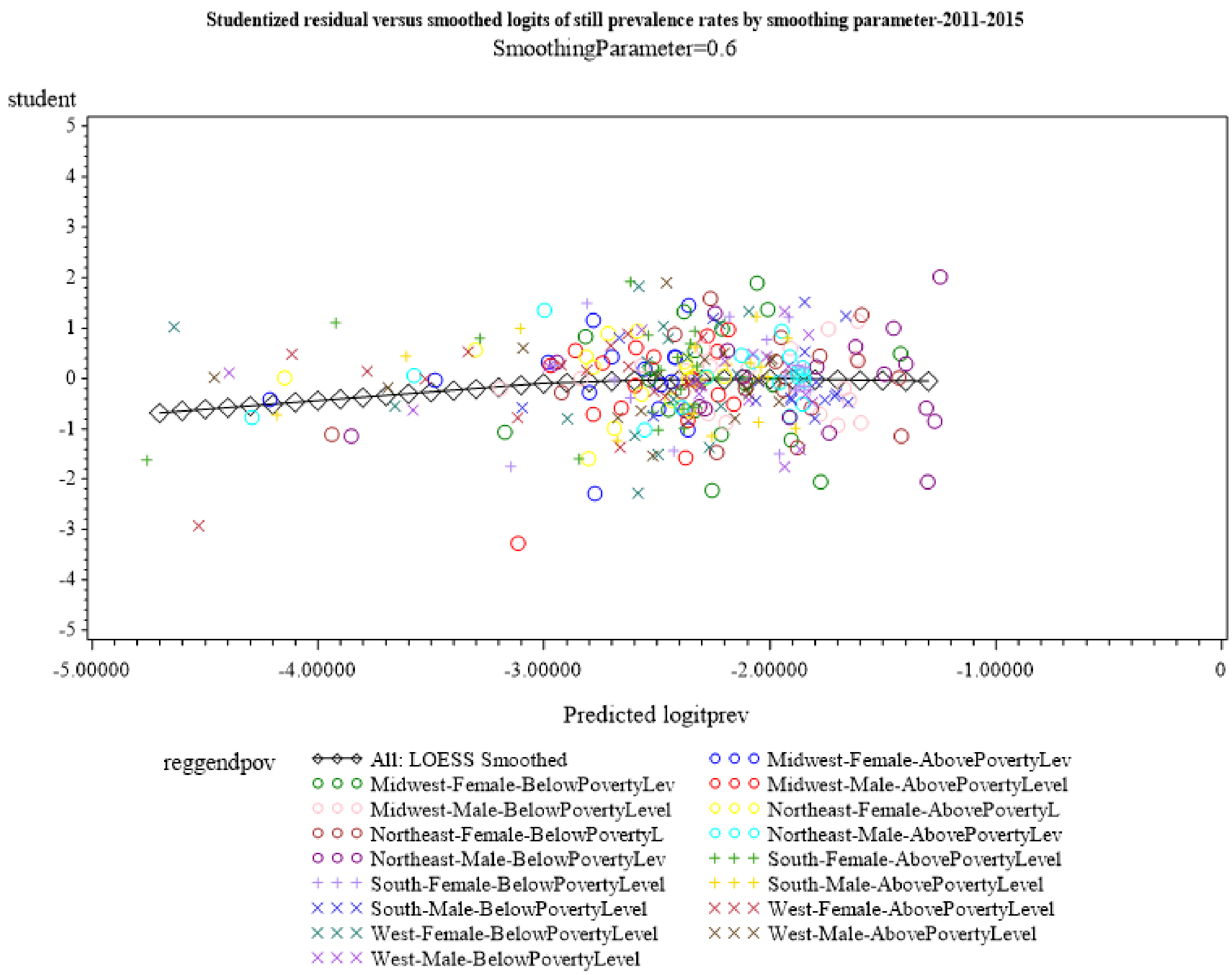


Figure E-2. Studentized residuals versus model predicted betas generated using a logistic model and using the children 'STILL' asthmatic data set, with smoothing set to 0.6.

When considering both children asthma prevalence responses evaluated, the residual standard error (estimated values for sigma) suggests the choice of smoothing parameter as varied, ranging from 0.4 to 0.7. The normal probability plots of the studentized residuals suggest preference for smoothing at or above 0.6. The plots of residuals against smoothed predictions suggest the choices of 0.4 through 0.6. We therefore chose the final value of 0.6 to use for smoothing the children's asthma prevalence. For the adults, there were small differences in the statistical metrics used to evaluate the smoothing. A value of 0.9 was selected for smoothing, consistent with what was used in my prior analysis (U.S. EPA, 2014).

The smoothed asthma prevalence and associated graphical presentation are provided in Attachment 2 following a similar format to that presented in Attachment 1.

Step 2: U.S. Census Tract Poverty Ratio Data Set Description and Processing

This section briefly describes the approach used to generate census tract level poverty ratios for all U.S. census tracts, stratified by age and age groups where available. Details regarding the data processing is provided below in Attachment 3.[8] Data used was from 2013 U.S. Census 5-year American Community Survey (ACS).

First, ACS internal point latitudes and longitudes were obtained from the 2013 Gazetteer files.[9] Next, the individual state level ACS sequence files (SF-56) were downloaded,[10] retaining the number of persons across the variable "B17024" for each state considering the appropriate logical record number.[11] The data provided by the B17024 variable is stratified by age or age groups (ages <5, 5, 6-11, 12-14, 15, 16-17, 18-24, 25-34, 35-44, 45-54, 55-64, 65-74, and ≥75) and income/poverty ratios, given in increments of 0.25. We calculated two new variables for each age using the number of persons from the B17024 stratifications; the fraction of those persons having poverty ratios < 1.5 and ≥ 1.5 by summing the appropriate B17024 variable and dividing by the total number of persons in that age/age group. Then, individual state level geographic data ("geo" files) and their associated documentation were downloaded[12] and

[8] Code has been adapted from ACS 2012 SAS programs and from ACS 2012 SAS Macros available at http://www2.census.gov/acs2012_5yr/summaryfile/UserTools/SF20125YR_SAS.zip and http://www2.census.gov/acs2012_5yr/summaryfile/UserTools/SF_All_Macro.sas

[9] Data set and content description is available at: http://www.census.gov/geo/maps-data/data/gazetteer2013.html.

[10] We used the summary tables (B17024), giving census tract populations by poverty income ratio and age group downloaded from http://www2.census.gov/acs2013_5yr/summaryfile/2009-2013_ACSSF_By_State_All_Tables/. We unzipped each state's ACS2013 5-yr table zip, then gathered sequence file 56.

[11] Information regarding variable names is available at https://www2.census.gov/acs2013_5yr/summaryfile/ACS_2013_SF_Tech_Doc.pdf. A file for the appropriate logical record number, "Sequence_Number_and_Table_Number_Lookup.xls", can be found at https://www2.census.gov/acs2013_5yr/summaryfile/.

[12] Geographic data were obtained from obtained from http://www2.census.gov/acs2013_5yr/summaryfile/2009-2013_ACSSF_By_State_All_Tables/b. Unzipped were each state's ACS2013 5-yr table ("g2013" file names).

screened for tract level information using the "sumlev" variable equal to '140'. Also identified was the US Region for each state, consistent with that used for the NHIS asthma prevalence data.[13]

Finally, the poverty ratio data were combined with the above described census tract level geographic data using the "stusab" and "logrecno" variables. Because APEX requires the input data files to be complete, additional processing of the poverty probability file was needed. For where there was missing tract level poverty information,[14] we substituted an age-specific value using the average for the particular county the tract was located within, or the state-wide average. The percent of tracts substituted using county averaged values varied by age group though, on average, was approximately 1.7% of the total tracts (Table E-4). Only a handful of tracts in six of the age groups were substituted using state averaged values.

Table E-4. Percent of tracts substituted with county average or state average poverty status.

Percent Substituted	Age Groups										
	≤5	6-11	12-17	18-24	25-34	35- 44	45-54	55-64	65-74	≥75	all
Filled with County Avg.	1.9	2.1	2.0	1.5	1.4	1.4	1.3	1.4	1.7	2.0	1.7
Filled with State Avg.	0.004	0.003	0.004	0.001	0	0	0	0	0	0.001	0.001

The final output was a single file containing relevant tract level poverty probabilities (pov_prob) by age groups for all U.S. census tracts.

Step 3: Combining Census Tract Poverty Ratios with the Asthma Prevalence Data

The two data sets were merged considering the region identifier and stratified by age and sex. The final asthma prevalence was calculated using the following weighting scheme:

*Asthma prevalence=round((pov_prob*prev_belowpov)+((**1**-pov_prob)*prev_abovepov),**0.0001**);*

whereas each U.S. census tract value now expresses a tract specific poverty-weighted asthma prevalence, stratified by ages (children 0-17), age groups (adults), and two sexes. These

[13] https://www2.census.gov/geo/pdfs/maps-data/maps/reference/us_regdiv.pdf

[14] Whether there were no data collected by the Census for the selected poverty status or whether there were simply no persons in that age group is relatively inconsequential to estimating the asthmatic persons exposed, particularly considering latter case as no persons in that age group would be modeled by APEX when using the same Census population data set.

final asthma prevalence data used for the assessment are found within the APEX *asthmaprevalence.txt* file.

REFERENCES

Cohen J and Rosenbaum A. (2005). Analysis of NHIS Asthma Prevalence Data. Memorandum to John Langstaff by ICF Incorporated. For US EPA Work Assignment 3-08 under EPA contract 68D01052. Available in US EPA (2007) Appendix G.

U.S. EPA. (2007). Ozone Population Exposure Analysis for Selected Urban Areas (July 2007). Office of Air Quality Planning and Standards, Research Triangle Park, NC. EPA-452/R-07-010. Available at: http://epa.gov/ttn/naaqs/standards/ozone/s_o3_cr_td.html.

U.S. EPA. (2008). Risk and Exposure Assessment to Support the Review of the NO_2 Primary National Ambient Air Quality Standard. Report no. EPA-452/R-08-008a. November 2008. Available at: http://www.epa.gov/ttn/naaqs/standards/nox/data/20081121_NO2_REA_final.pdf.

U.S. EPA. (2009). Risk and Exposure Assessment to Support the Review of the SO_2 Primary National Ambient Air Quality Standard. Report no. EPA-452/R-09-007. August 2009. Available at: http://www.epa.gov/ttn/naaqs/standards/so2/data/200908SO2REAFinalReport.pdf.

U.S. EPA. (2014). Health Risk and Exposure Assessment for Ozone, Final Report. Chapter 5 Appendices. Report no. EPA-452/R-14-004c. August 2014. Available at https://nepis.epa.gov/Exe/ZyPDF.cgi/P100KCI7.PDF?Dockey=P100KCI7.PDF.

Attachment 1 – Non-Smoothed Asthma Prevalence (Figures 1 - 4)

Figure 1 - Children (Ever Have Asthma)

Figure 1. Raw asthma 'EVER' prevalence rates and confidence intervals-2011-2015
region=Midwest pov_rat=Above Poverty Level

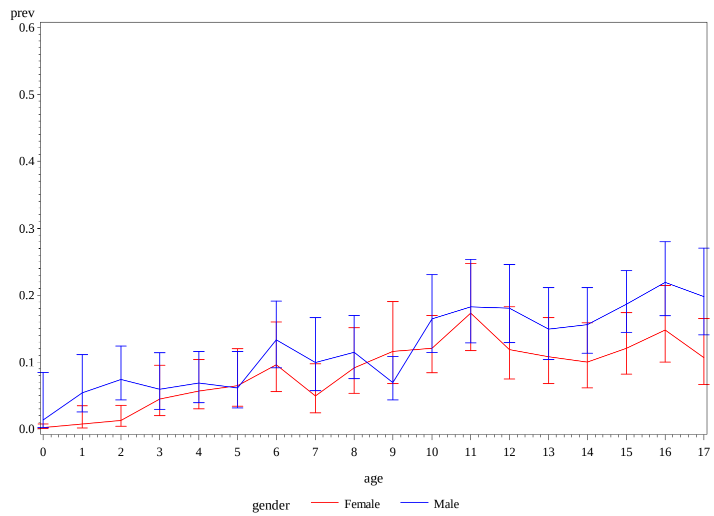


Figure 1. Raw asthma 'EVER' prevalence rates and confidence intervals-2011-2015
region=Midwest pov_rat=Below Poverty Level

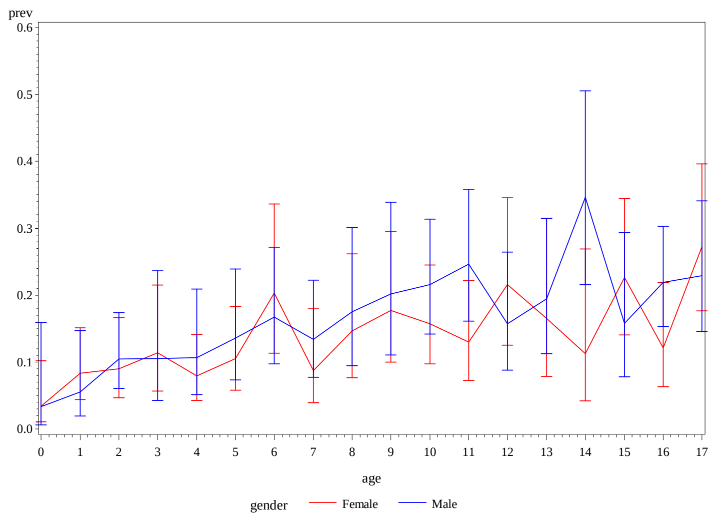


Figure 1. Raw asthma 'EVER' prevalence rates and confidence intervals-2011-2015
region=Northeast pov_rat=Above Poverty Level

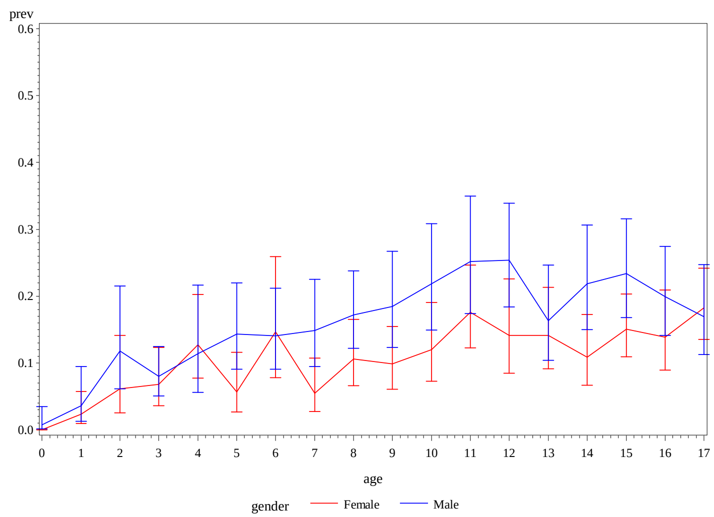


Figure 1. Raw asthma 'EVER' prevalence rates and confidence intervals-2011-2015
region=Northeast pov_rat=Below Poverty Level

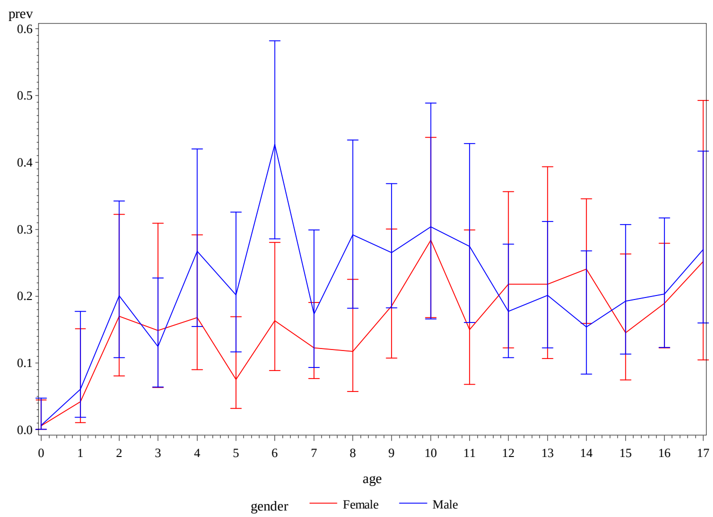


Figure 1, cont. - Children (Ever Have Asthma)

Figure 1. Raw asthma 'EVER' prevalence rates and confidence intervals-2011-2015
region=South pov_rat=Above Poverty Level

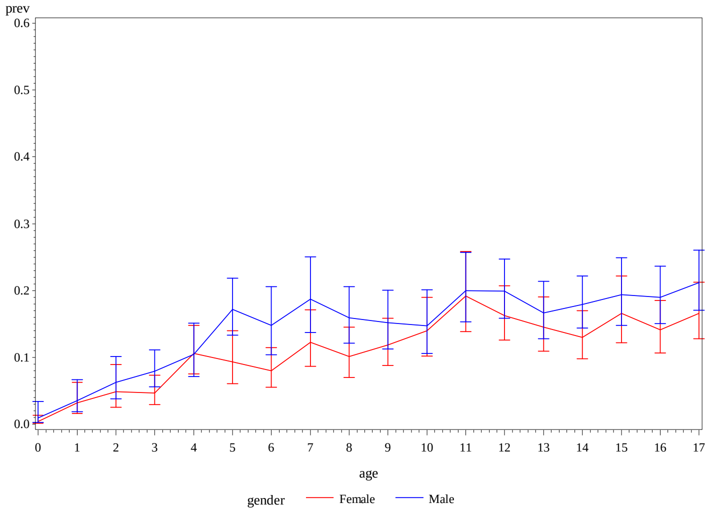


Figure 1. Raw asthma 'EVER' prevalence rates and confidence intervals-2011-2015
region=South pov_rat=Below Poverty Level

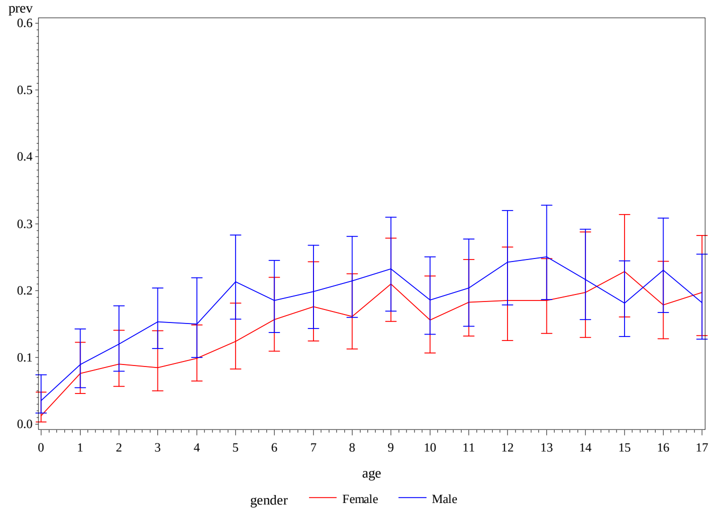


Figure 1. Raw asthma 'EVER' prevalence rates and confidence intervals-2011-2015
region=West pov_rat=Above Poverty Level

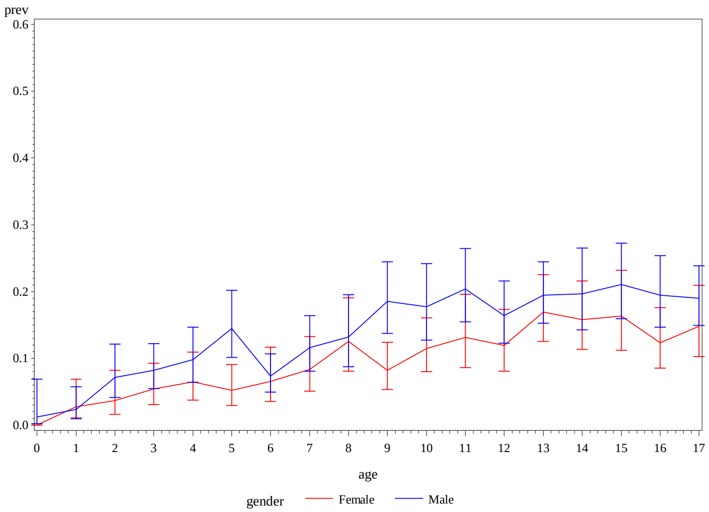


Figure 1. Raw asthma 'EVER' prevalence rates and confidence intervals-2011-2015
region=West pov_rat=Below Poverty Level

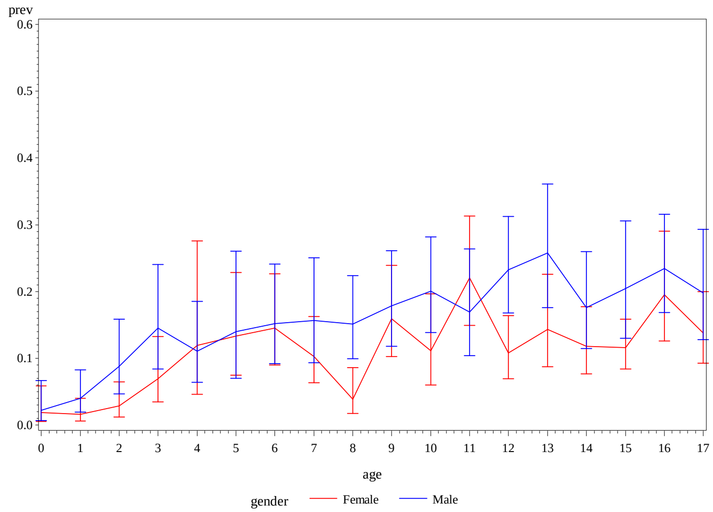


Figure 2 – Children (Still Have Asthma)

Figure 2. Raw asthma 'STILL' prevalence rates and confidence intervals-2011-2015
region=Midwest pov_rat=Above Poverty Level

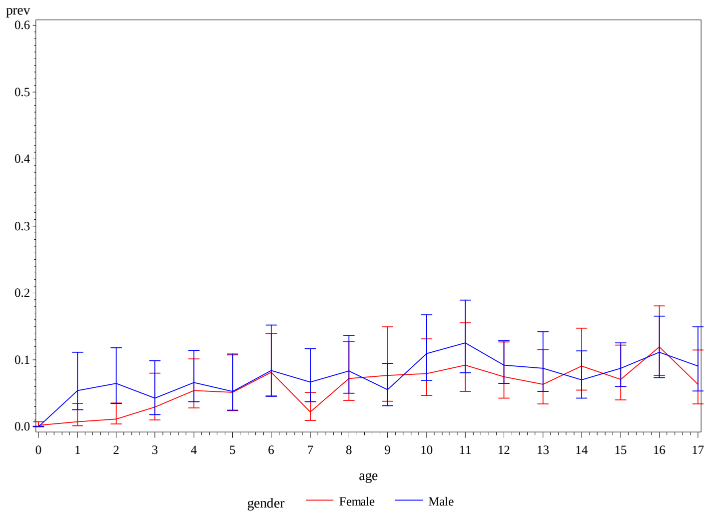


Figure 2. Raw asthma 'STILL' prevalence rates and confidence intervals-2011-2015
region=Midwest pov_rat=Below Poverty Level

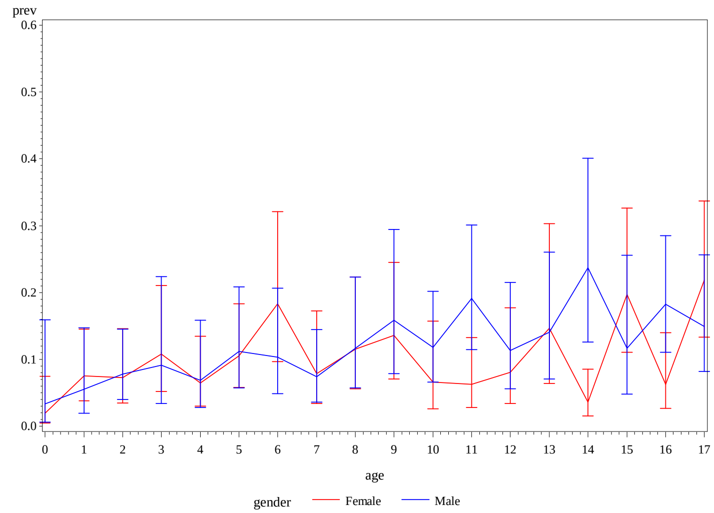


Figure 2. Raw asthma 'STILL' prevalence rates and confidence intervals-2011-2015
region=Northeast pov_rat=Above Poverty Level

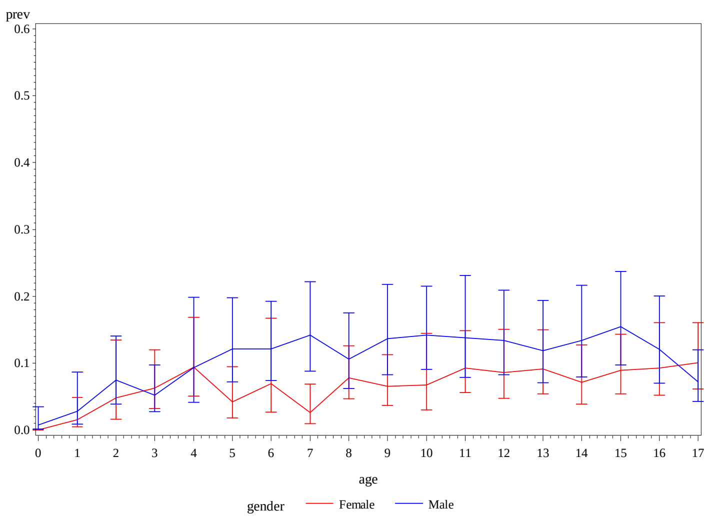


Figure 2. Raw asthma 'STILL' prevalence rates and confidence intervals-2011-2015
region=Northeast pov_rat=Below Poverty Level

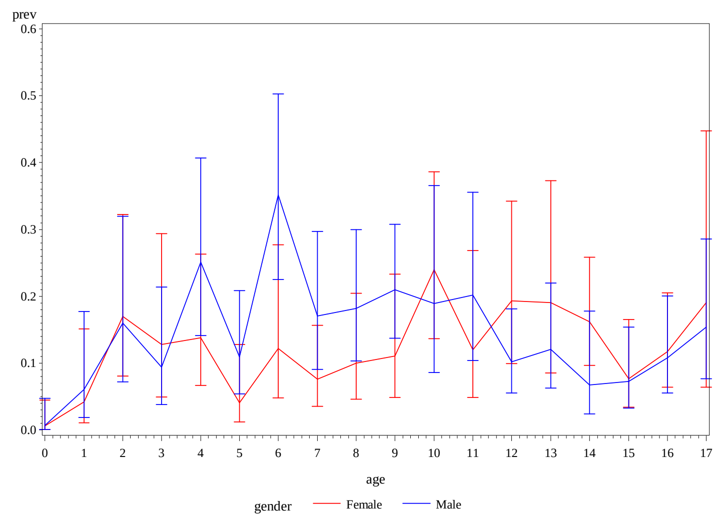


Figure 2, cont. – Children (Still Have Asthma)

Figure 2. Raw asthma 'STILL' prevalence rates and confidence intervals-2011-2015
region=South pov_rat=Above Poverty Level

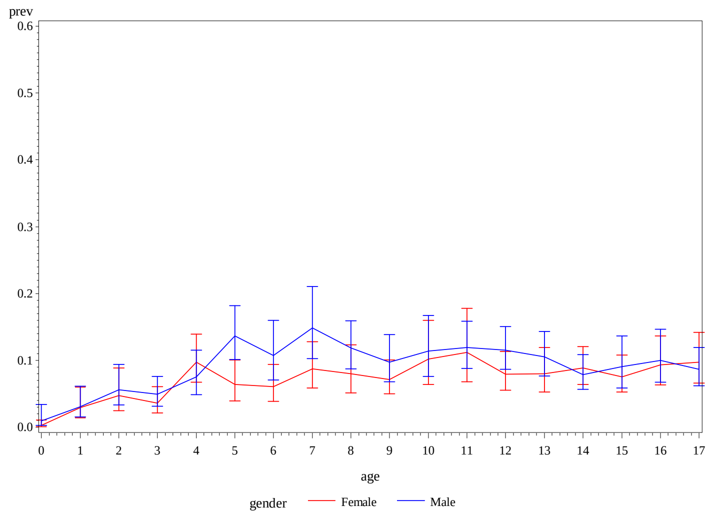


Figure 2. Raw asthma 'STILL' prevalence rates and confidence intervals-2011-2015
region=South pov_rat=Below Poverty Level

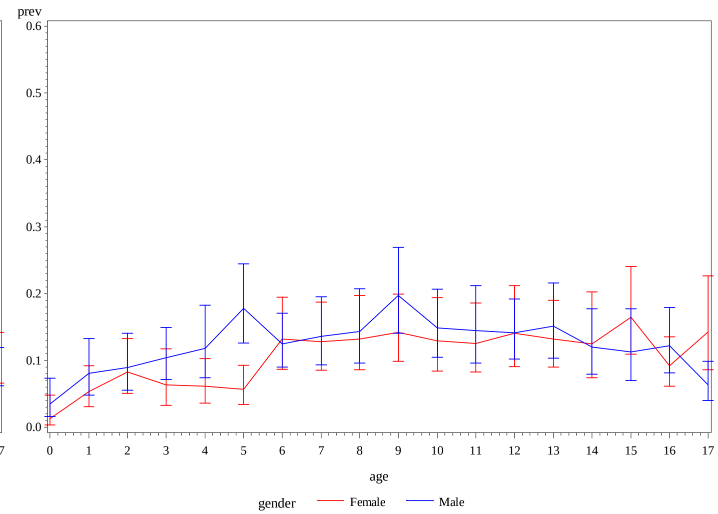


Figure 2. Raw asthma 'STILL' prevalence rates and confidence intervals-2011-2015
region=West pov_rat=Above Poverty Level

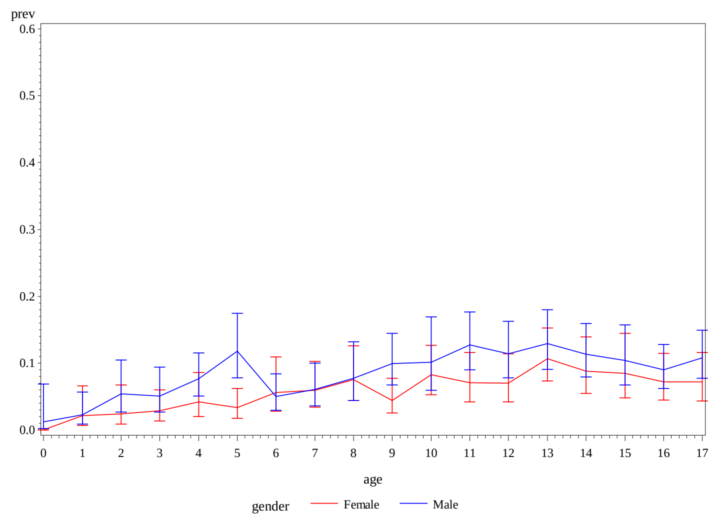


Figure 2. Raw asthma 'STILL' prevalence rates and confidence intervals-2011-2015
region=West pov_rat=Below Poverty Level

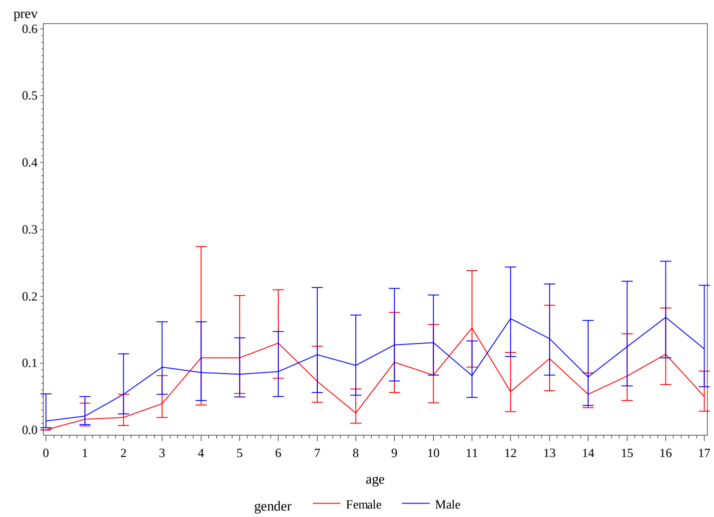
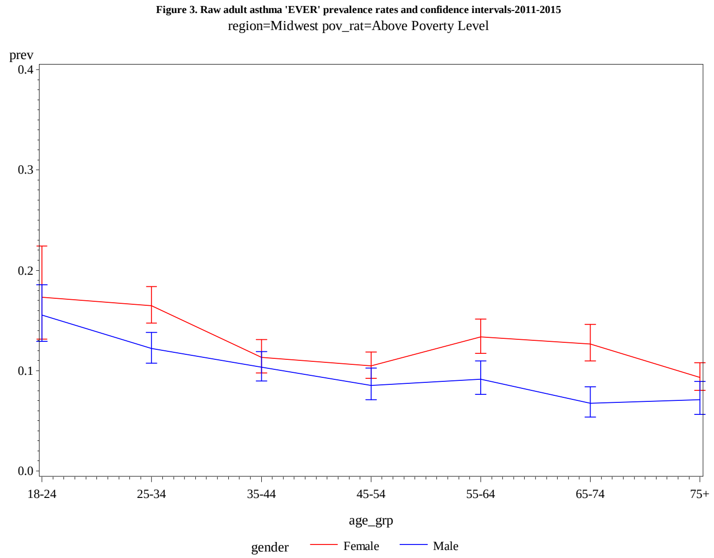
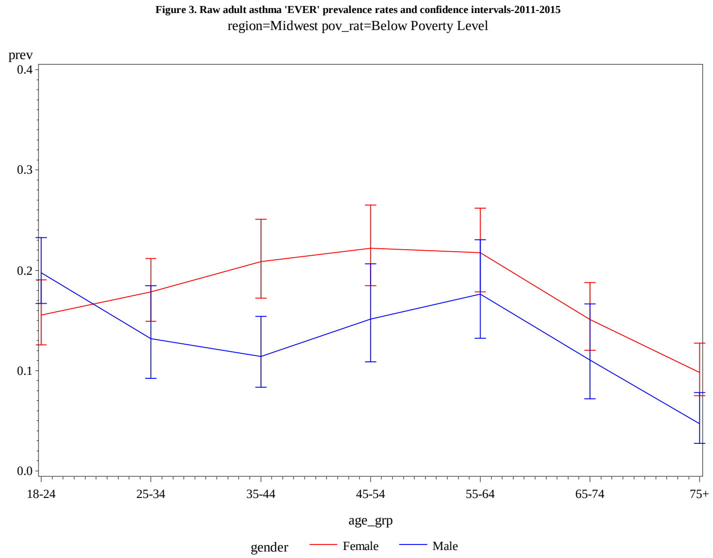
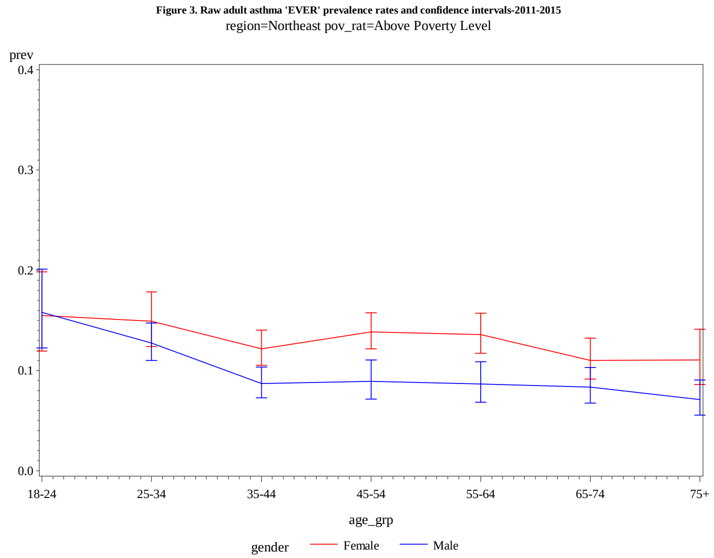


Figure 3 – Adults (Ever Have Asthma)







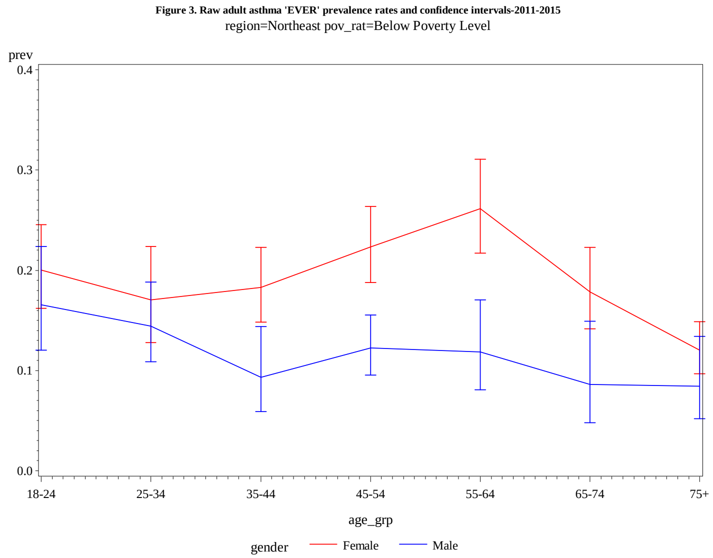
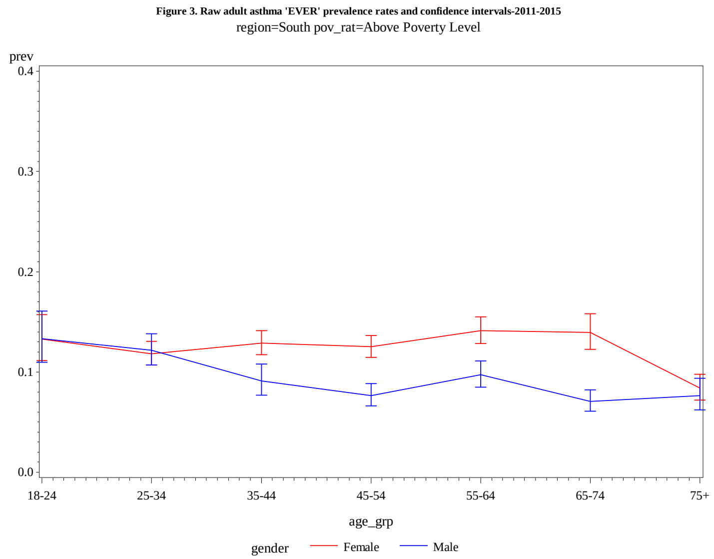
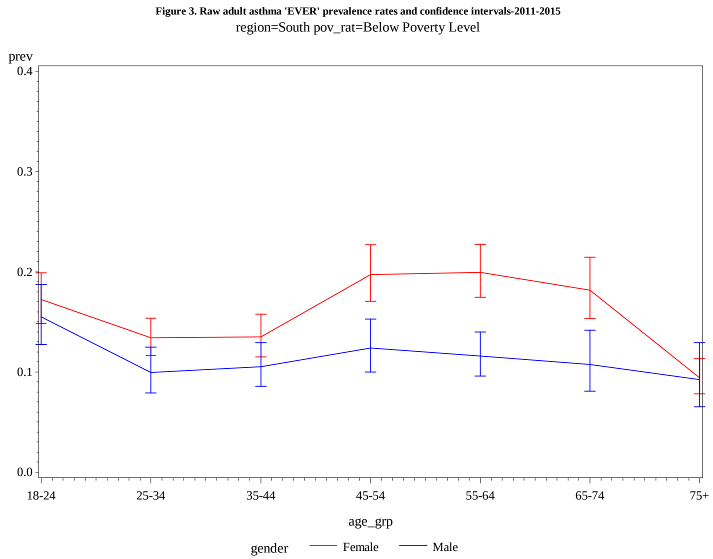
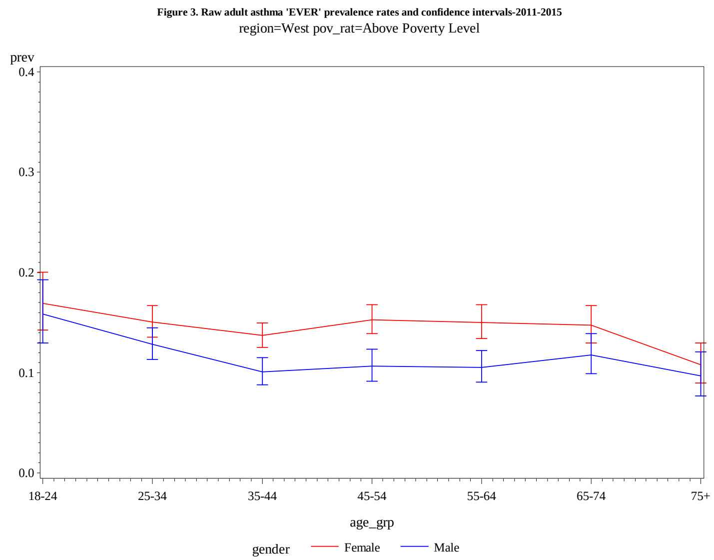


Figure 3, cont. – Adults (Ever Have Asthma)







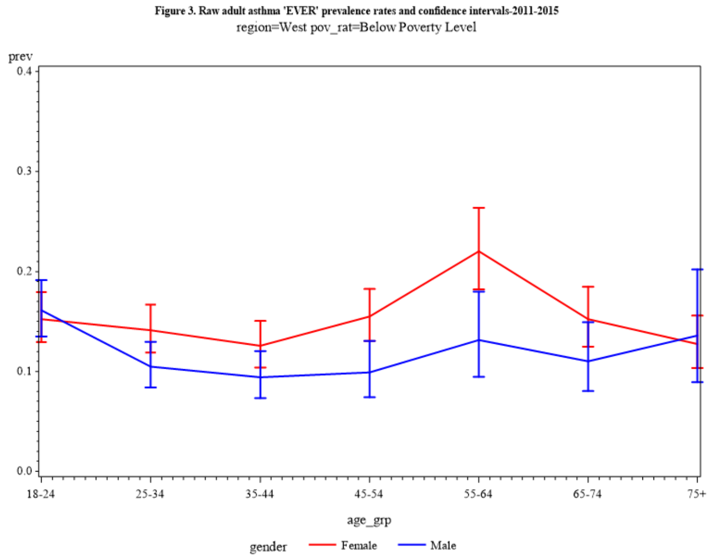


Figure 4 – Adults (Still Have Asthma)

Figure 4. Raw adult asthma 'STILL' prevalence rates and confidence intervals-2011-2015
region=Midwest pov_rat=Above Poverty Level

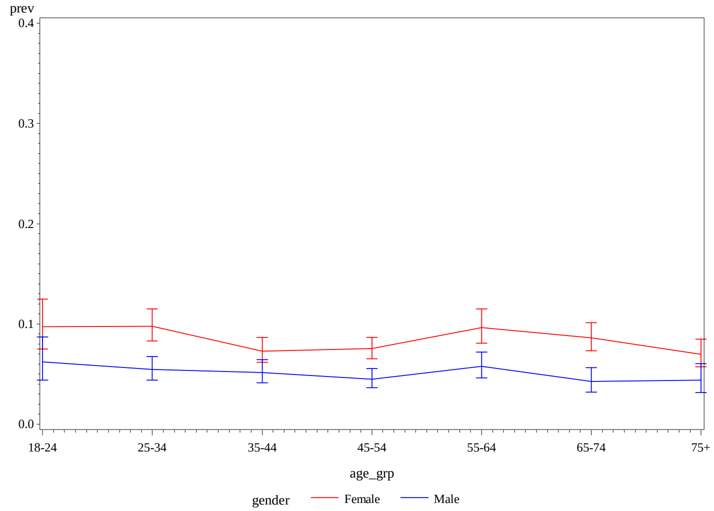


Figure 4. Raw adult asthma 'STILL' prevalence rates and confidence intervals-2011-2015
region=Midwest pov_rat=Below Poverty Level

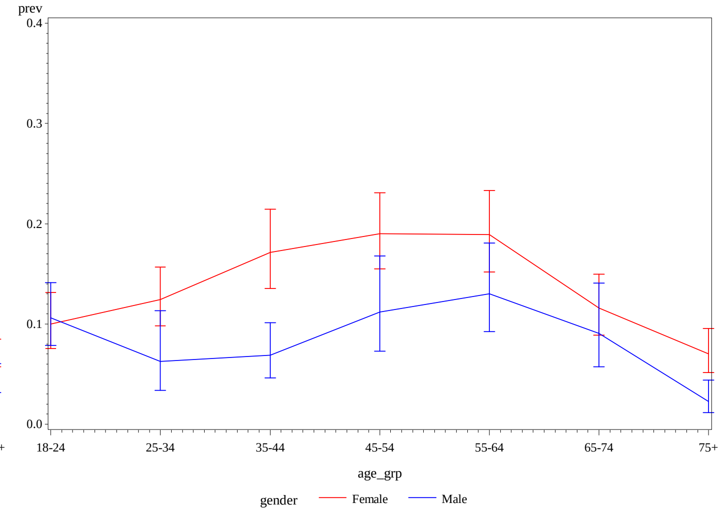


Figure 4. Raw adult asthma 'STILL' prevalence rates and confidence intervals-2011-2015
region=Northeast pov_rat=Above Poverty Level

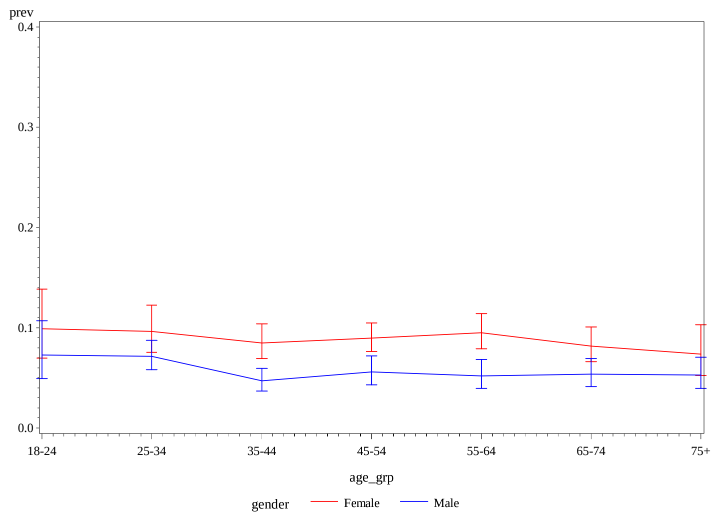


Figure 4. Raw adult asthma 'STILL' prevalence rates and confidence intervals-2011-2015
region=Northeast pov_rat=Below Poverty Level

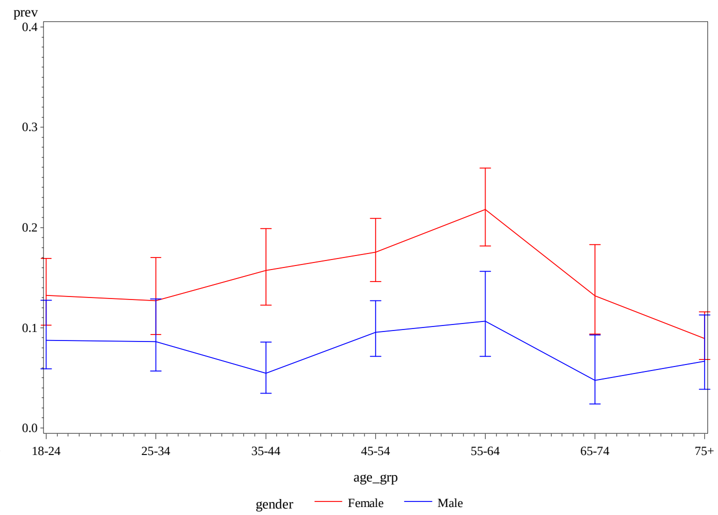


Figure 4, cont. – Adults (Still Have Asthma)

Figure 4. Raw adult asthma 'STILL' prevalence rates and confidence intervals-2011-2015
region=South pov_rat=Above Poverty Level

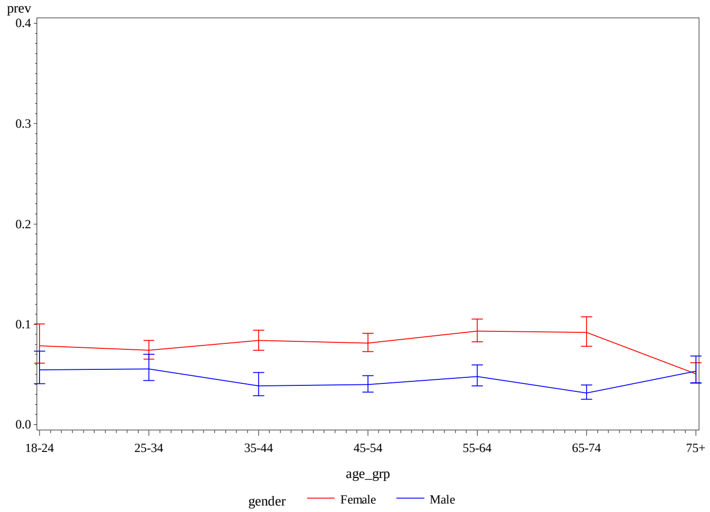


Figure 4. Raw adult asthma 'STILL' prevalence rates and confidence intervals-2011-2015
region=South pov_rat=Below Poverty Level

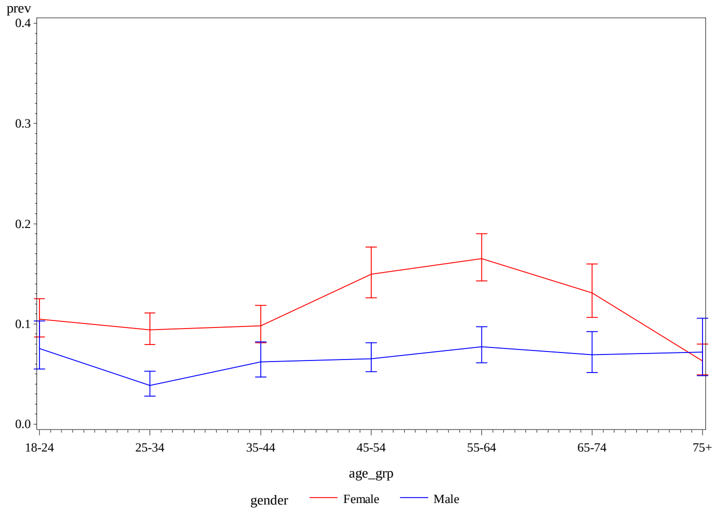


Figure 4. Raw adult asthma 'STILL' prevalence rates and confidence intervals-2011-2015
region=West pov_rat=Below Poverty Level

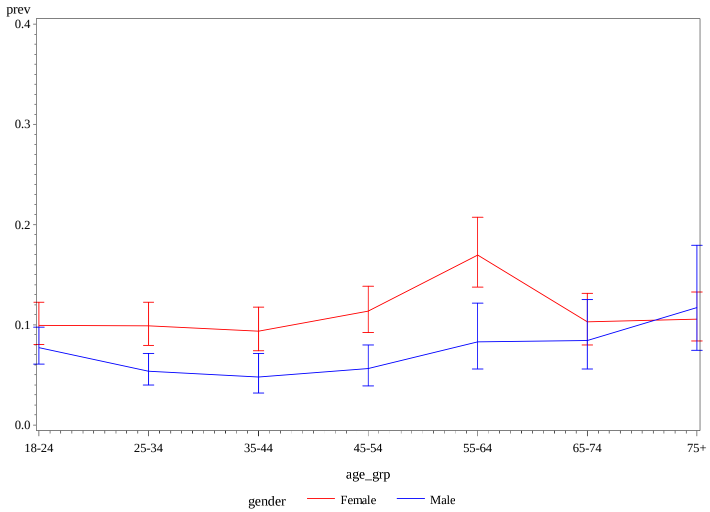
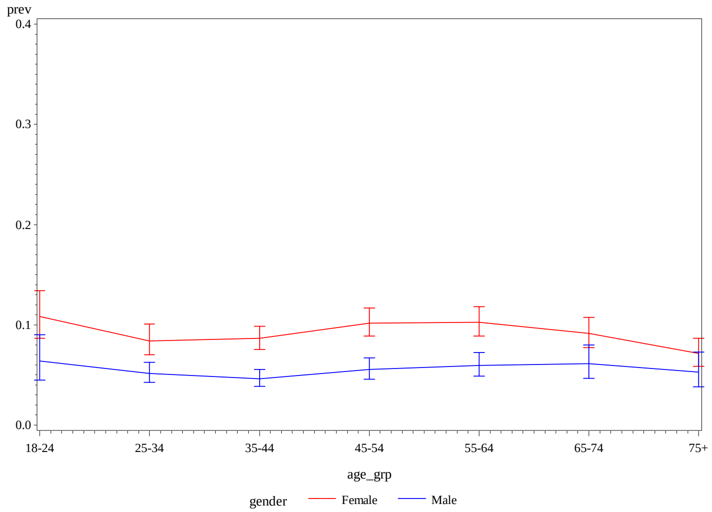


Figure 4. Raw adult asthma 'STILL' prevalence rates and confidence intervals-2011-2015
region=West pov_rat=Above Poverty Level



Attachment 2 –Smoothed Asthma Prevalence (Figures 1-4)

Figure 1 – Children (Ever Have Asthma)

Figure 1. Smoothed asthma 'EVER' prevalence rates and confidence intervals-2011-2015
region=Midwest pov_rat=Above Poverty Level

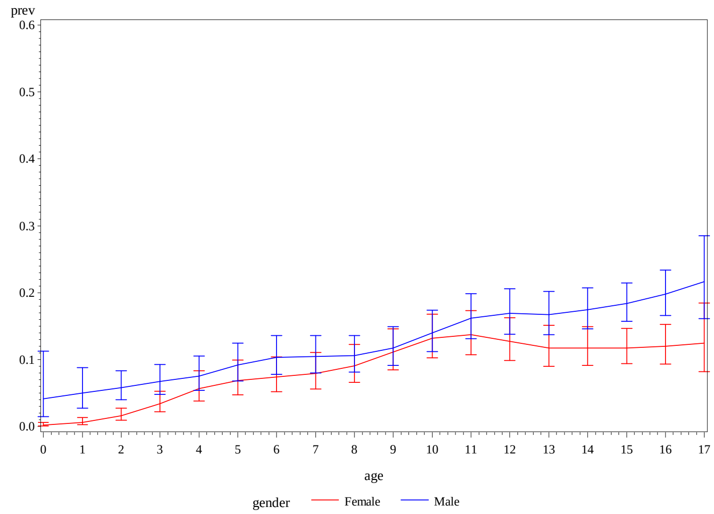


Figure 1. Smoothed asthma 'EVER' prevalence rates and confidence intervals-2011-2015
region=Midwest pov_rat=Below Poverty Level

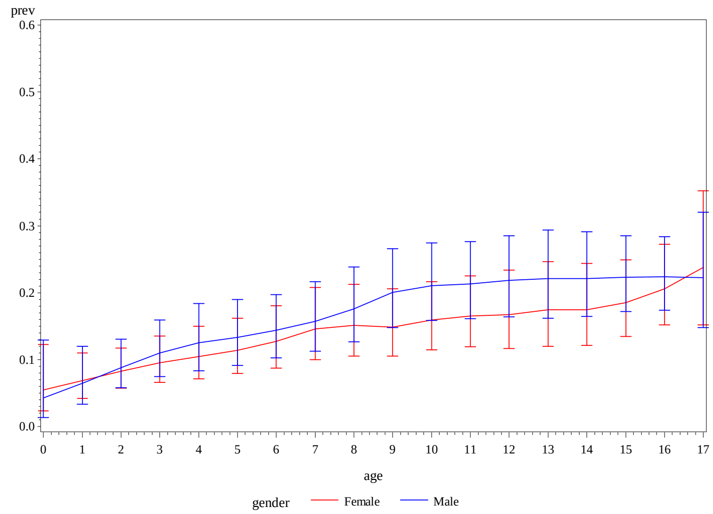


Figure 1. Smoothed asthma 'EVER' prevalence rates and confidence intervals-2011-2015
region=Northeast pov_rat=Above Poverty Level

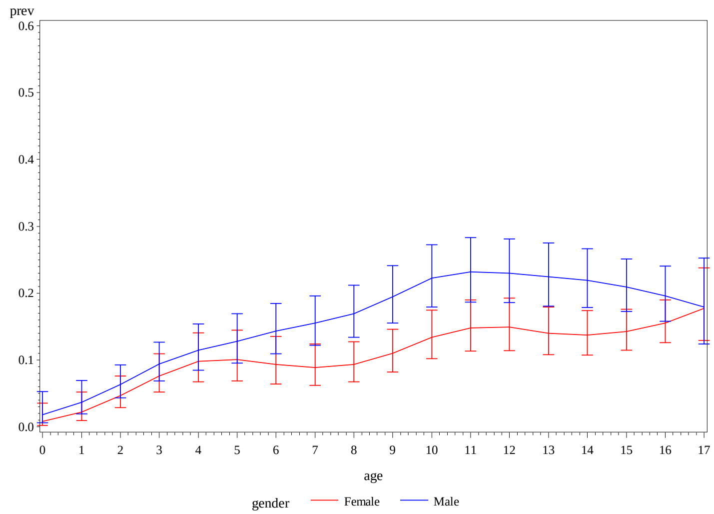


Figure 1. Smoothed asthma 'EVER' prevalence rates and confidence intervals-2011-2015
region=Northeast pov_rat=Below Poverty Level

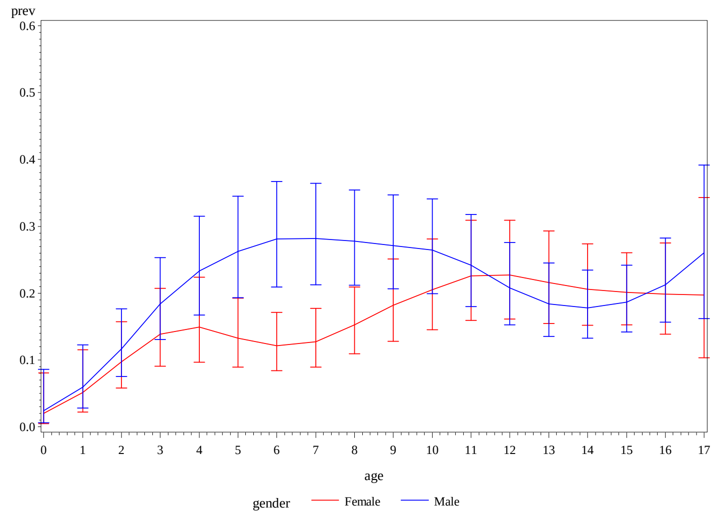


Figure 1, cont. – Children (Ever Have Asthma)

Figure 1. Smoothed asthma 'EVER' prevalence rates and confidence intervals-2011-2015
region=South pov_rat=Above Poverty Level

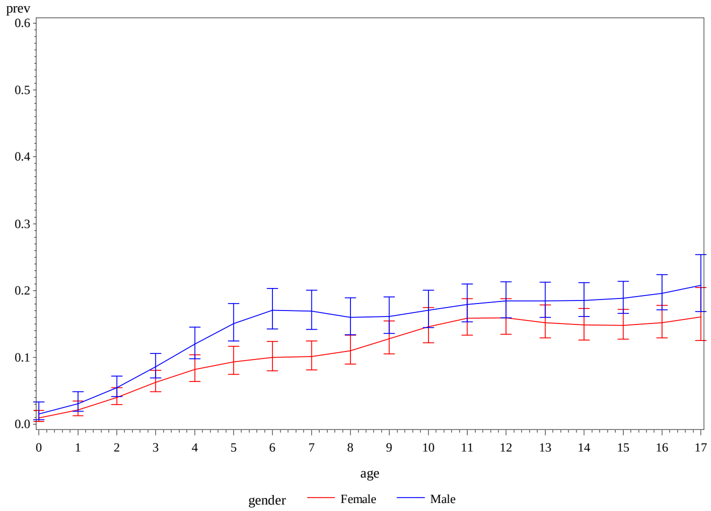


Figure 1. Smoothed asthma 'EVER' prevalence rates and confidence intervals-2011-2015
region=South pov_rat=Below Poverty Level

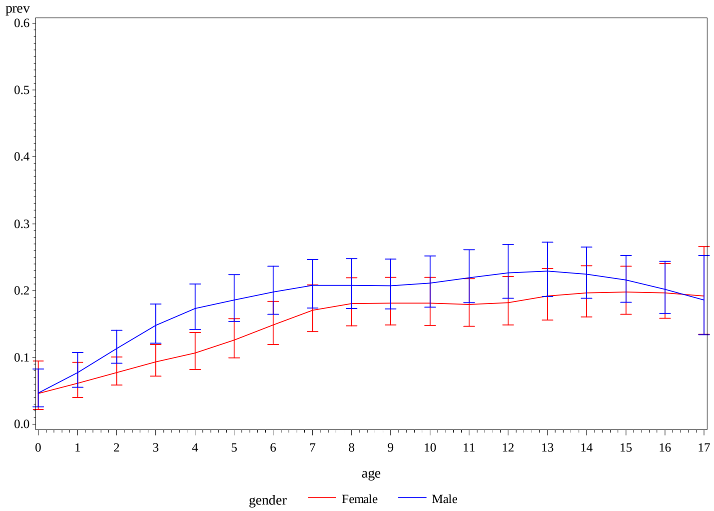


Figure 1. Smoothed asthma 'EVER' prevalence rates and confidence intervals-2011-2015
region=West pov_rat=Above Poverty Level

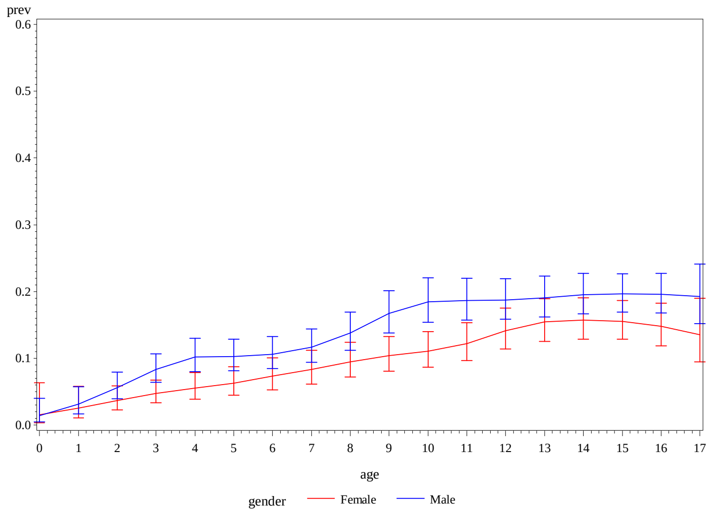


Figure 1. Smoothed asthma 'EVER' prevalence rates and confidence intervals-2011-2015
region=West pov_rat=Below Poverty Level

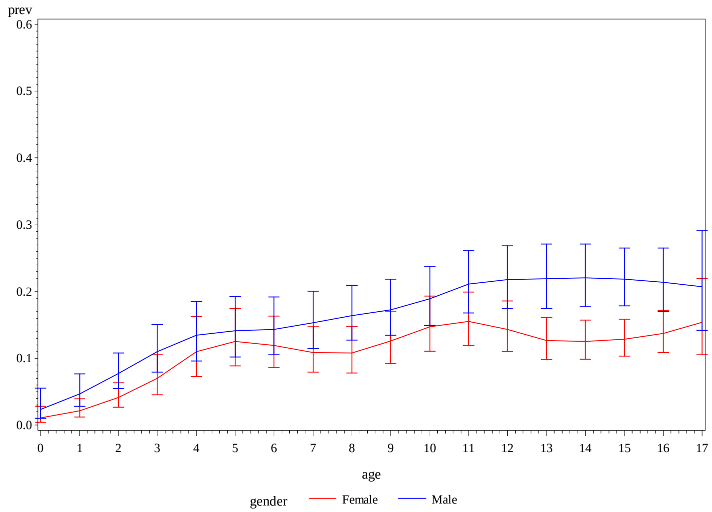


Figure 2 – Children (Still Have Asthma)

Figure 2. Smoothed asthma 'STILL' prevalence rates and confidence intervals-2011-2015
region=Midwest pov_rat=Above Poverty Level

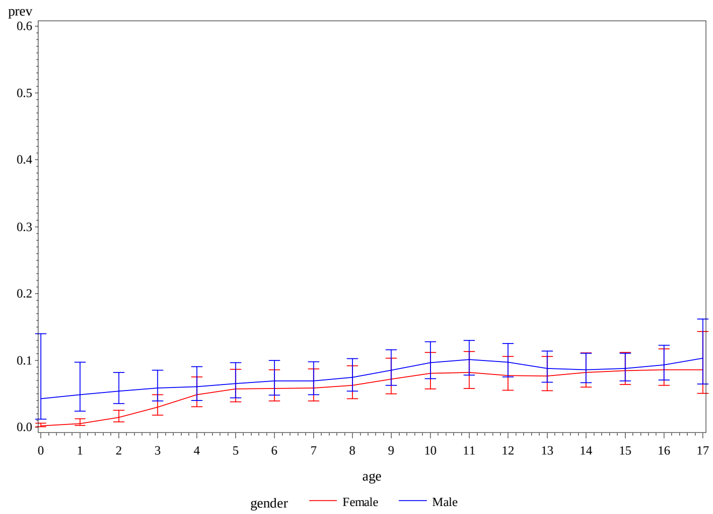


Figure 2. Smoothed asthma 'STILL' prevalence rates and confidence intervals-2011-2015
region=Midwest pov_rat=Below Poverty Level

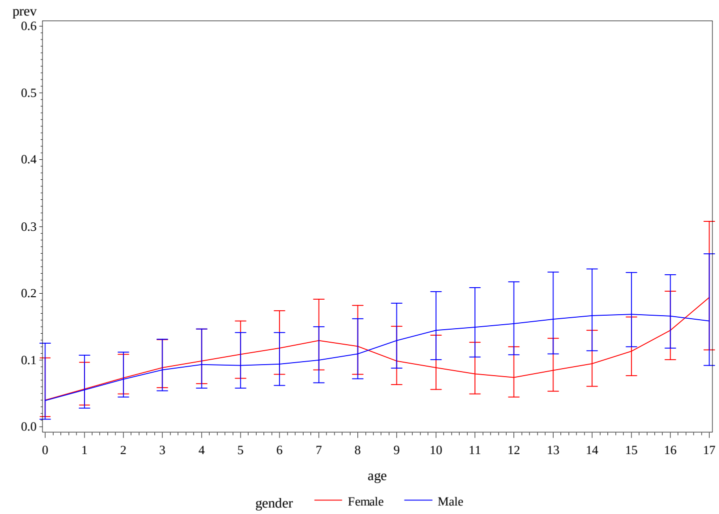


Figure 2. Smoothed asthma 'STILL' prevalence rates and confidence intervals-2011-2015
region=Northeast pov_rat=Above Poverty Level

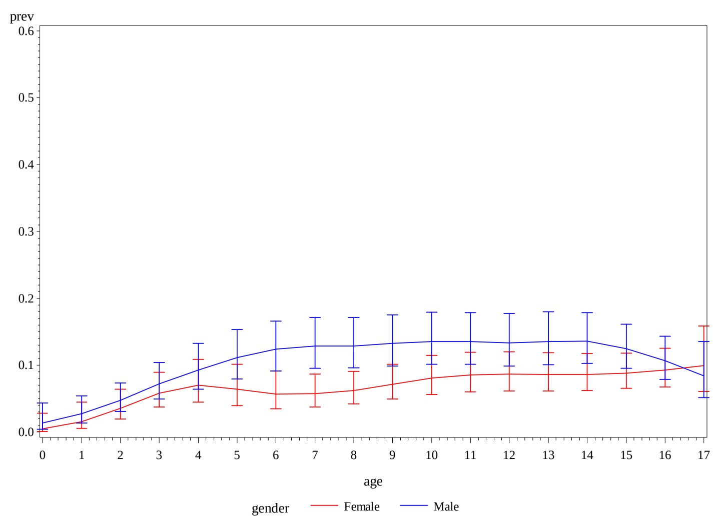


Figure 2. Smoothed asthma 'STILL' prevalence rates and confidence intervals-2011-2015
region=Northeast pov_rat=Below Poverty Level

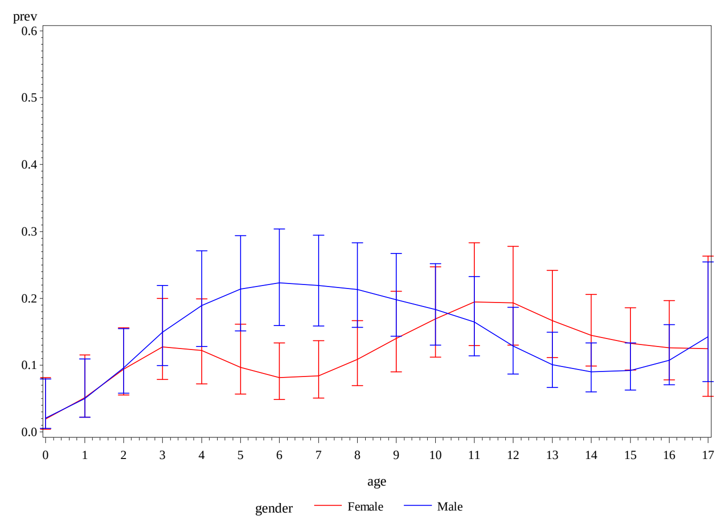


Figure 2, cont. – Children (Still Have Asthma)

Figure 2. Smoothed asthma 'STILL' prevalence rates and confidence intervals-2011-2015
region=South pov_rat=Above Poverty Level

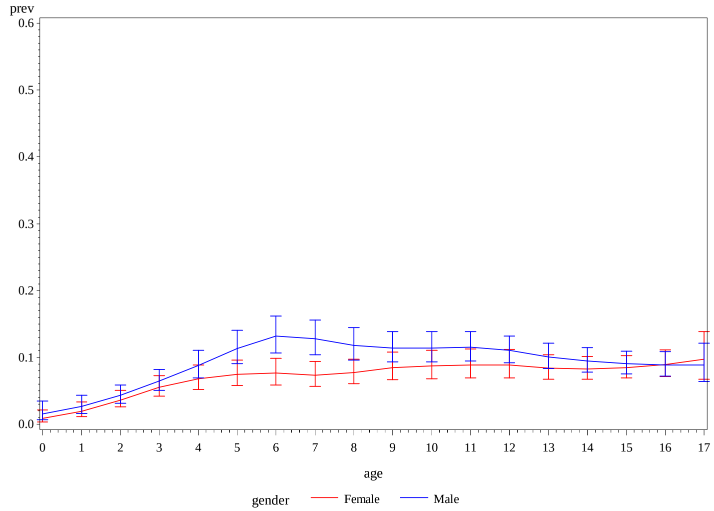


Figure 2. Smoothed asthma 'STILL' prevalence rates and confidence intervals-2011-2015
region=South pov_rat=Below Poverty Level

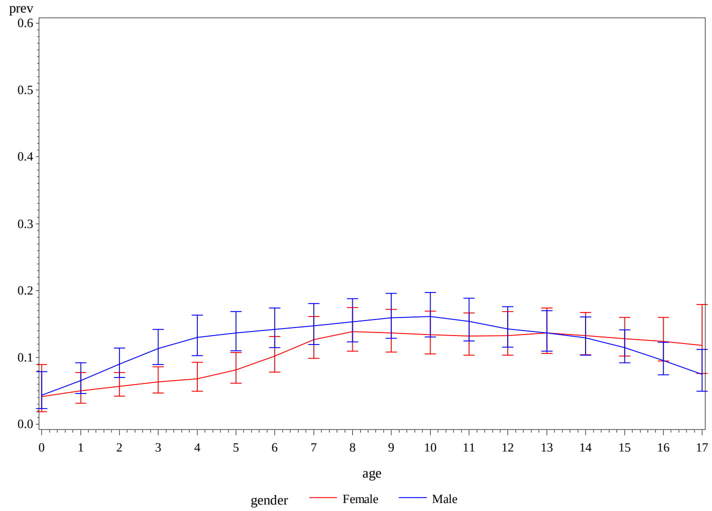


Figure 2. Smoothed asthma 'STILL' prevalence rates and confidence intervals-2011-2015
region=West pov_rat=Above Poverty Level

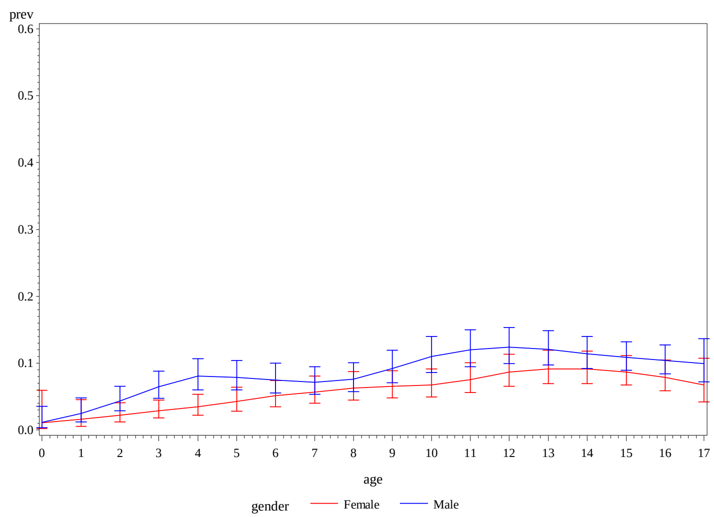


Figure 2. Smoothed asthma 'STILL' prevalence rates and confidence intervals-2011-2015
region=West pov_rat=Below Poverty Level

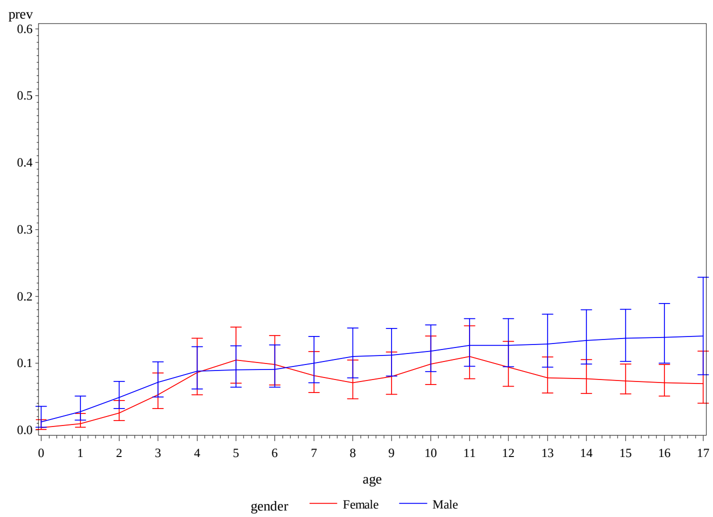
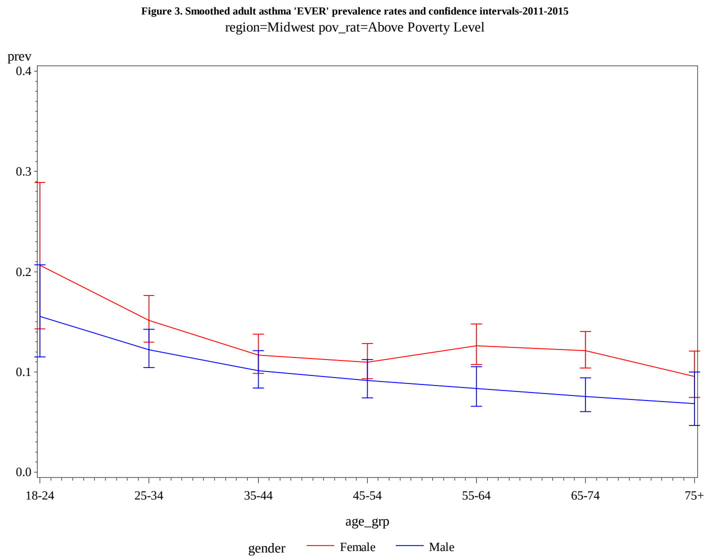
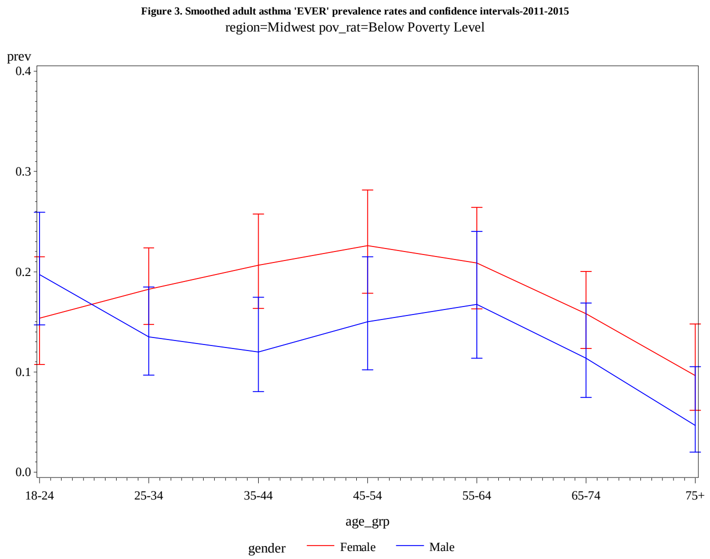
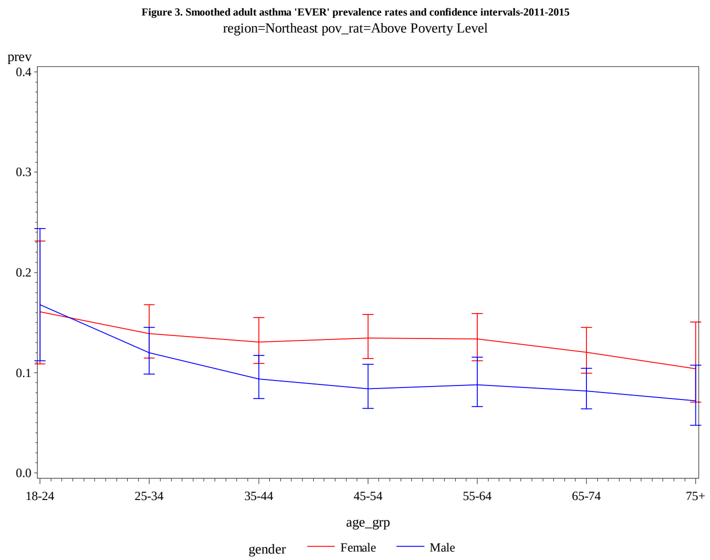


Figure 3 – Adults (Ever Have Asthma)







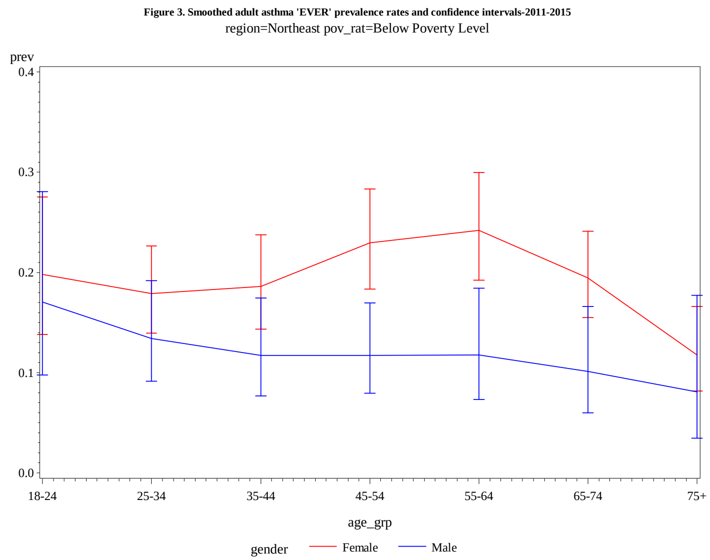


Figure 3, cont. – Adults (Ever Have Asthma)

Figure 3. Smoothed adult asthma 'EVER' prevalence rates and confidence intervals-2011-2015
region=South pov_rat=Above Poverty Level

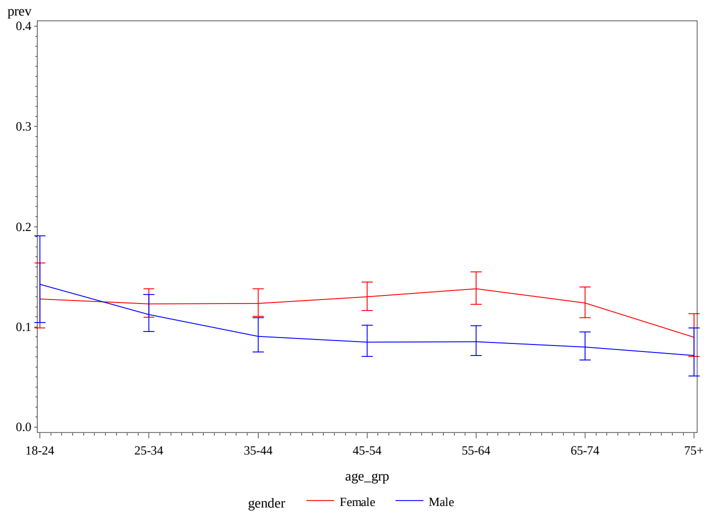


Figure 3. Smoothed adult asthma 'EVER' prevalence rates and confidence intervals-2011-2015
region=South pov_rat=Below Poverty Level

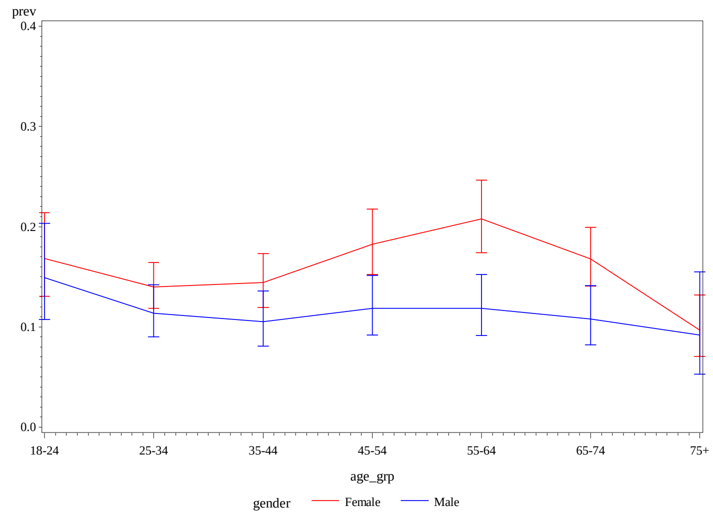


Figure 3. Smoothed adult asthma 'EVER' prevalence rates and confidence intervals-2011-2015
region=West pov_rat=Above Poverty Level

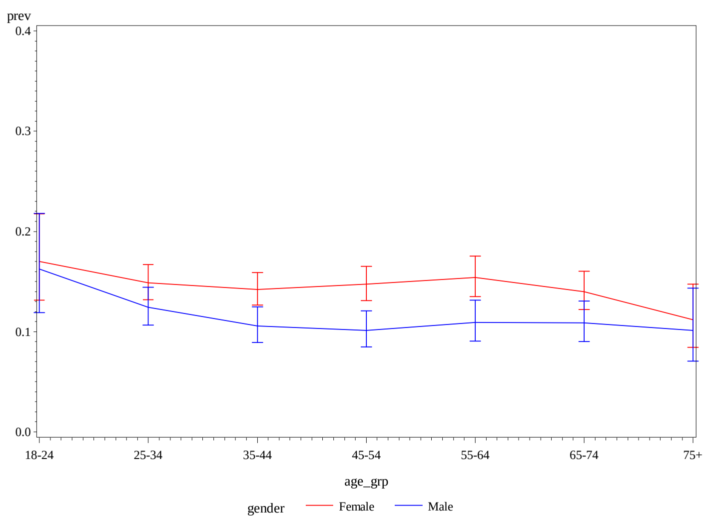


Figure 3. Smoothed adult asthma 'EVER' prevalence rates and confidence intervals-2011-2015
region=West pov_rat=Below Poverty Level

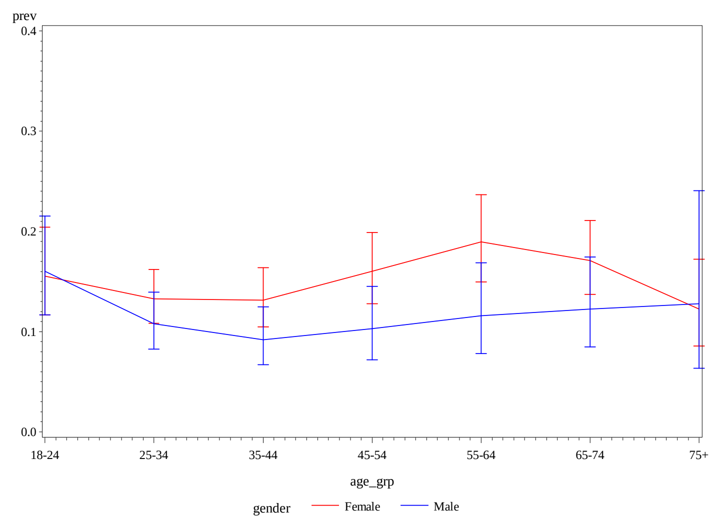


Figure 4 – Adults (Still Have Asthma)

Figure 4. Smoothed adult asthma 'STILL' prevalence rates and confidence intervals-2011-2015
region=Midwest pov_rat=Above Poverty Level

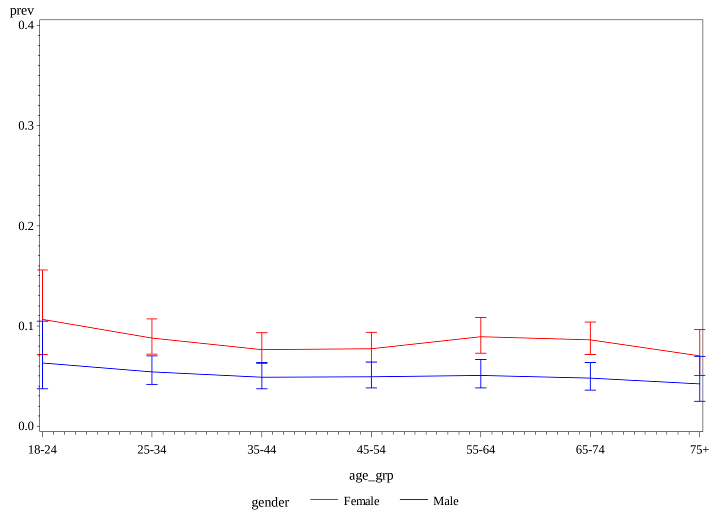


Figure 4. Smoothed adult asthma 'STILL' prevalence rates and confidence intervals-2011-2015
region=Midwest pov_rat=Below Poverty Level

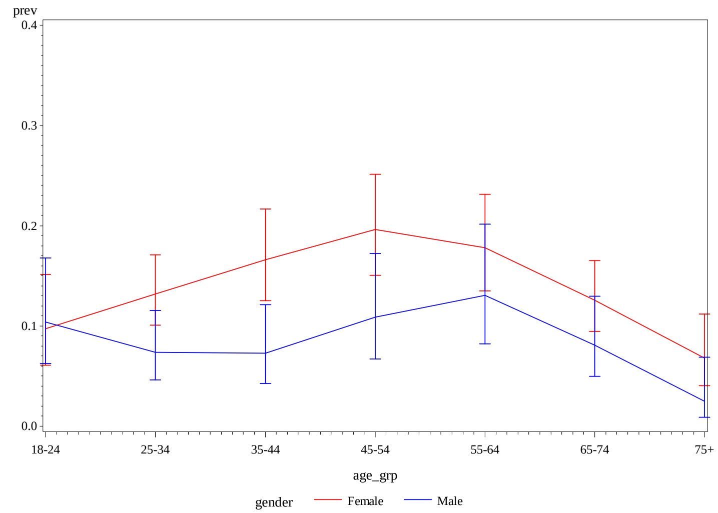


Figure 4. Smoothed adult asthma 'STILL' prevalence rates and confidence intervals-2011-2015
region=Northeast pov_rat=Above Poverty Level

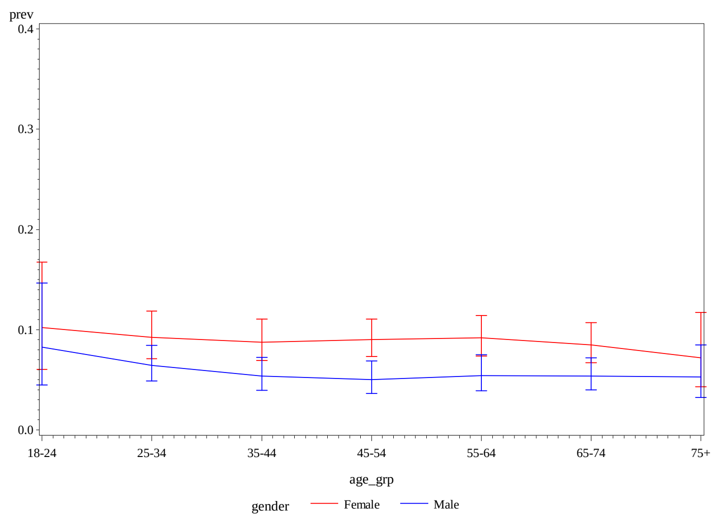


Figure 4. Smoothed adult asthma 'STILL' prevalence rates and confidence intervals-2011-2015
region=Northeast pov_rat=Below Poverty Level

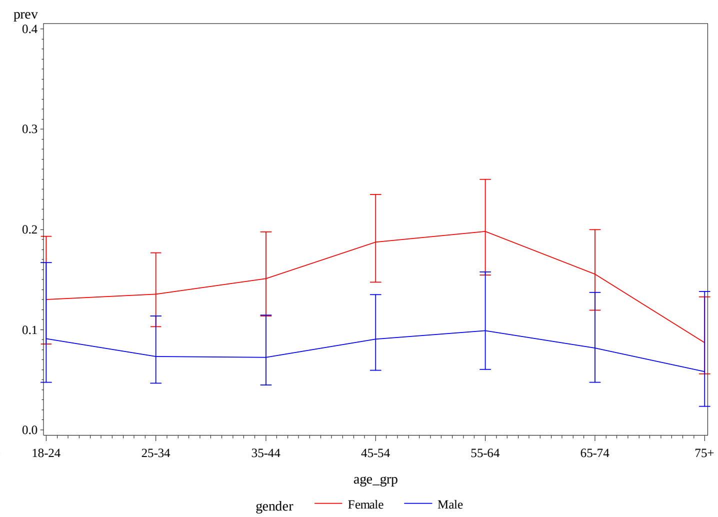


Figure 4, cont. – Adults (Still Have Asthma)

Figure 4. Smoothed adult asthma 'STILL' prevalence rates and confidence intervals-2011-2015
region=South pov_rat=Above Poverty Level

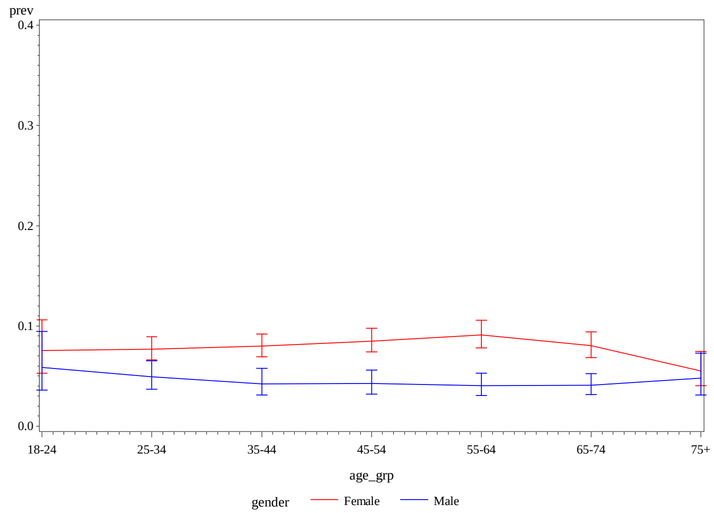


Figure 4. Smoothed adult asthma 'STILL' prevalence rates and confidence intervals-2011-2015
region=South pov_rat=Below Poverty Level

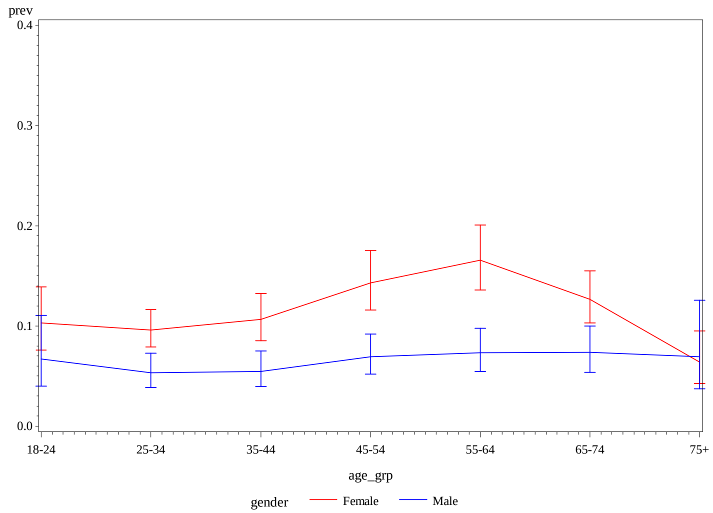


Figure 4. Smoothed adult asthma 'STILL' prevalence rates and confidence intervals-2011-2015
region=West pov_rat=Above Poverty Level

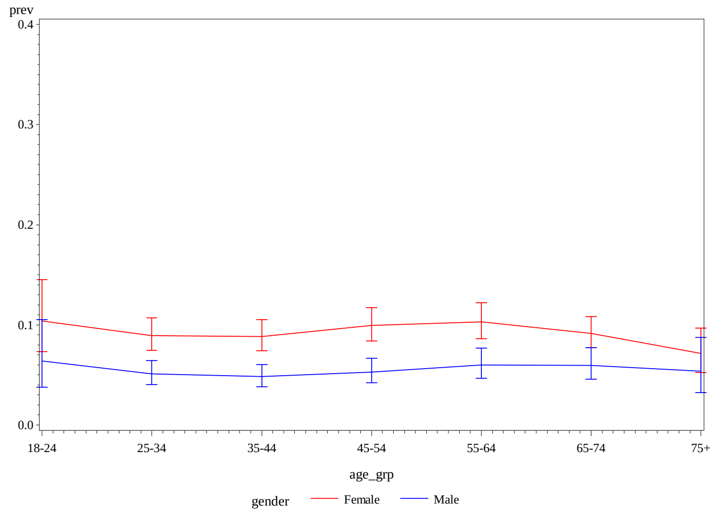
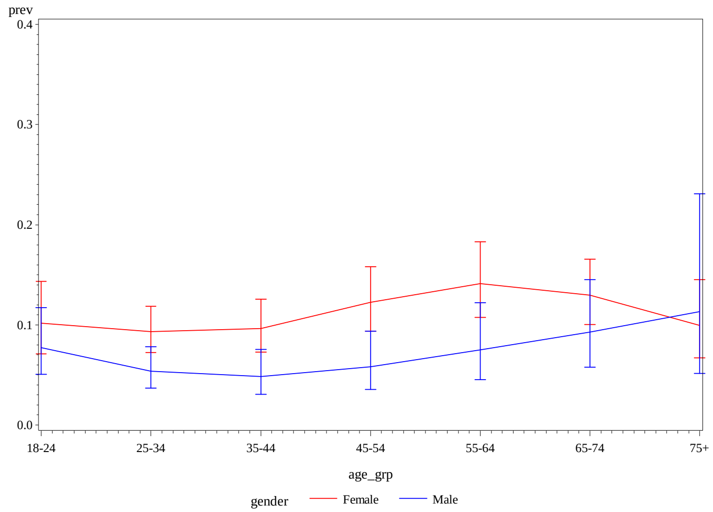


Figure 4. Smoothed adult asthma 'STILL' prevalence rates and confidence intervals-2011-2015
region=West pov_rat=Below Poverty Level



Attachment 3 – Processing Code for US Census Poverty Status Data from 2013 ACS

```
options mlogic;

LIBNAME sas 'F:\SGRAHAM\NHIS\NHIS_1115_Process'; run; *location of sas data library;

*imports ACS2013_5yr internal point Latitude and Longitude;
PROC IMPORT OUT= acs2013_5yr_tract_lat_long
        DATAFILE= "F:\SGRAHAM\NHIS\NHIS_1115_Process\2013_Gaz_tracts_national.txt"
        DBMS=TAB REPLACE;
    GETNAMES=YES;
    DATAROW=2;
RUN;

*formats a new variable GEOID_merge using LAT LONs GEOID in order to merge LAT and LON to geography dataset by GEOID_merge;
data sas.acs2013_5yr_tract_lat_long (keep = GEOID_merge LAT LON);
        set work.acs2013_5yr_tract_lat_long(rename=(GEOID=GEOID_char INTPTLAT=LAT INTPTLONG=LON));
        length GEOID_merge $12.;
        GEOID_merge = put(GEOID_char,Best12.); *STATE COUNTY and TRACT from ACS2013 Sequence File data make up GEOID in Lat Lon file;
run;

%macro Read_poverty(geo); *Imports ACS2013_5yr sequence file 56, income/poverty data (Table B17024) by state (geo);
DATA work.SFe0056&geo;
        LENGTH FILEID   $6
                FILETYPE $6
                STUSAB   $2
                CHARITER $3
                SEQUENCE $4
                LOGRECNO $7;

INFILE "F:\SGRAHAM\NHIS\NHIS_1115_Process\e20135&geo.0056000.txt" DSD TRUNCOVER DELIMITER =',' LRECL=3000;

LABEL
        FILEID  ='File Identification'
        FILETYPE='File Type'
        STUSAB  ='State/U.S.-Abbreviation (USPS)'
        CHARITER='Character Iteration'
        SEQUENCE='Sequence Number'
        LOGRECNO='Logical Record Number'

        /*AGE BY RATIO OF INCOME TO POVERTY LEVEL IN THE PAST 12 MONTHS */
        /*Universe: Population for whom poverty status is determined */

        B17024e1='Total:'
        B17024e2='Under 6 years:'
        B17024e3='Under .50'
        B17024e4='.50 to .74'
        B17024e5='.75 to .99'
        B17024e6='1.00 to 1.24'
        B17024e7='1.25 to 1.49'
        B17024e8='1.50 to 1.74'
        B17024e9='1.75 to 1.84'
        B17024e10='1.85 to 1.99'
        B17024e11='2.00 to 2.99'
        B17024e12='3.00 to 3.99'
        B17024e13='4.00 to 4.99'
        B17024e14='5.00 and over'
        B17024e15='6 to 11 years:'
        B17024e16='Under .50'
        B17024e17='.50 to .74'
        B17024e18='.75 to .99'
        B17024e19='1.00 to 1.24'
        B17024e20='1.25 to 1.49'
        B17024e21='1.50 to 1.74'
        B17024e22='1.75 to 1.84'
        B17024e23='1.85 to 1.99'
        B17024e24='2.00 to 2.99'
        B17024e25='3.00 to 3.99'
        B17024e26='4.00 to 4.99'
        B17024e27='5.00 and over'
        B17024e28='12 to 17 years:'
        B17024e29='Under .50'
        B17024e30='.50 to .74'
        B17024e31='.75 to .99'
        B17024e32='1.00 to 1.24'
        B17024e33='1.25 to 1.49'
        B17024e34='1.50 to 1.74'
        B17024e35='1.75 to 1.84'
        B17024e36='1.85 to 1.99'
        B17024e37='2.00 to 2.99'
        B17024e38='3.00 to 3.99'
        B17024e39='4.00 to 4.99'
        B17024e40='5.00 and over'
        B17024e41='18 to 24 years:'
```

B17024e42='Under .50'
B17024e43='.50 to .74'
B17024e44='.75 to .99'
B17024e45='1.00 to 1.24'
B17024e46='1.25 to 1.49'
B17024e47='1.50 to 1.74'
B17024e48='1.75 to 1.84'
B17024e49='1.85 to 1.99'
B17024e50='2.00 to 2.99'
B17024e51='3.00 to 3.99'
B17024e52='4.00 to 4.99'
B17024e53='5.00 and over'
B17024e54='25 to 34 years:'
B17024e55='Under .50'
B17024e56='.50 to .74'
B17024e57='.75 to .99'
B17024e58='1.00 to 1.24'
B17024e59='1.25 to 1.49'
B17024e60='1.50 to 1.74'
B17024e61='1.75 to 1.84'
B17024e62='1.85 to 1.99'
B17024e63='2.00 to 2.99'
B17024e64='3.00 to 3.99'
B17024e65='4.00 to 4.99'
B17024e66='5.00 and over'
B17024e67='35 to 44 years:'
B17024e68='Under .50'
B17024e69='.50 to .74'
B17024e70='.75 to .99'
B17024e71='1.00 to 1.24'
B17024e72='1.25 to 1.49'
B17024e73='1.50 to 1.74'
B17024e74='1.75 to 1.84'
B17024e75='1.85 to 1.99'
B17024e76='2.00 to 2.99'
B17024e77='3.00 to 3.99'
B17024e78='4.00 to 4.99'
B17024e79='5.00 and over'
B17024e80='45 to 54 years:'
B17024e81='Under .50'
B17024e82='.50 to .74'
B17024e83='.75 to .99'
B17024e84='1.00 to 1.24'
B17024e85='1.25 to 1.49'
B17024e86='1.50 to 1.74'
B17024e87='1.75 to 1.84'
B17024e88='1.85 to 1.99'
B17024e89='2.00 to 2.99'
B17024e90='3.00 to 3.99'
B17024e91='4.00 to 4.99'
B17024e92='5.00 and over'
B17024e93='55 to 64 years:'
B17024e94='Under .50'
B17024e95='.50 to .74'
B17024e96='.75 to .99'
B17024e97='1.00 to 1.24'
B17024e98='1.25 to 1.49'
B17024e99='1.50 to 1.74'
B17024e100='1.75 to 1.84'
B17024e101='1.85 to 1.99'
B17024e102='2.00 to 2.99'
B17024e103='3.00 to 3.99'
B17024e104='4.00 to 4.99'
B17024e105='5.00 and over'
B17024e106='65 to 74 years:'
B17024e107='Under .50'
B17024e108='.50 to .74'
B17024e109='.75 to .99'
B17024e110='1.00 to 1.24'
B17024e111='1.25 to 1.49'
B17024e112='1.50 to 1.74'
B17024e113='1.75 to 1.84'
B17024e114='1.85 to 1.99'
B17024e115='2.00 to 2.99'
B17024e116='3.00 to 3.99'
B17024e117='4.00 to 4.99'
B17024e118='5.00 and over'
B17024e119='75 years and over:'
B17024e120='Under .50'
B17024e121='.50 to .74'
B17024e122='.75 to .99'
B17024e123='1.00 to 1.24'
B17024e124='1.25 to 1.49'
B17024e125='1.50 to 1.74'
B17024e126='1.75 to 1.84'
B17024e127='1.85 to 1.99'
B17024e128='2.00 to 2.99'
B17024e129='3.00 to 3.99'

```
                B17024e130='4.00 to 4.99'
                B17024e131='5.00 and over'
                ;

INPUT
                FILEID  $
                FILETYPE $
                STUSAB  $
                CHARITER $
                SEQUENCE $
                LOGRECNO $
                B17024e1-B17024e131
                ;
                if B17024e1 >=0;
RUN;
%mend;

%macro AnyGeo(geo); *Imports geo data file, assigns a census region, limits to 2013ACS_5yr census tracts by state ('geo'), assigns lat lon;
data work.g20135&geo (drop =          AIANHH                    AIANHHFP     AIHHTLI      AITS          AITSCE                ANRC        BLKGRP       CBSA
                CDCURR        CNECTA
                                                                                                        COMPONENT  CONCIT                  COUSUB       CSA
                DIVISION       FILEID          MACC          MEMI          METDIV        NAME
                                                                                                        NECTA                     NECTADIV     PCI
                PLACE          PUMA1                         PUMA5         REGION        SDELM          SDSEC         SDUNI
                                                                                                        SLDL                      SLDU                        STATECE
                SUBMCD         SUMLEVEL        TAZ                         UA                           UACP          UGA                       UR
                                                                                                        US                                     VTD
                ZCTA3          ZCTA5
                                               );

/*Location of geo data file for import*/
 INFILE "F:\SGRAHAM\NHIS\NHIS_1115_Process\g20135&geo..txt" MISSOVER TRUNCOVER LRECL=500; /*change directory*/

 LABEL FILEID  ='File Identification'          STUSAB  ='State Postal Abbreviation'
                        SUMLEVEL='Summary Level'                   COMPONENT='geographic Component'
                        LOGRECNO='Logical Record Number'            US     ='US'
                        REGION  ='Region'                                                                   DIVISION ='Division'
                        STATECE ='State (Census Code)'                                STATE    ='State (FIPS Code)'
                        COUNTY  ='County'                                                                   COUSUB  ='County Subdivision (FIPS)'
                        PLACE   ='Place (FIPS Code)'                        TRACT   ='Census Tract'
                        BLKGRP  ='Block Group'                                                   CONCIT   ='Consolidated City'
                        CSA     ='Combined Statistical Area'       METDIV  ='Metropolitan Division'
                        UA      ='Urban Area'           UACP    ='Urban Area Central Place'
                        VTD     ='Voting District'                                       ZCTA3  ='ZIP Code Tabulation Area (3-digit)'
                        SUBMCD  ='Subbarrio (FIPS)'                                      SDELM  ='School District (Elementary)'
                        SDSEC   ='School District (Secondary)'      SDUNI  ='School District (Unified)'
                        UR      ='Urban/Rural'                                                   PCI    ='Principal City Indicator'
                        TAZ     ='Traffic Analysis Zone'            UGA    ='Urban Growth Area'
                        GEOID   ='geographic Identifier'                    NAME   ='Area Name'
                        AIANHH  ='American Indian Area/Alaska Native Area/Hawaiian Home Land (Census)'
                        AIANHHFP='American Indian Area/Alaska Native Area/Hawaiian Home Land (FIPS)'
                        AIHHTLI ='American Indian Trust Land/Hawaiian Home Land Indicator'
                        AITSCE  ='American Indian Tribal Subdivision (Census)'
                        AITS    ='American Indian Tribal Subdivision (FIPS)'
                        ANRC    ='Alaska Native Regional Corporation (FIPS)'
                        CBSA    ='Metropolitan and Micropolitan Statistical Area'
                        MACC    ='Metropolitan Area Central City'
                        MEMI    ='Metropolitan/Micropolitan Indicator Flag'
                        NECTA   ='New England City and Town Combined Statistical Area'
                        CNECTA  ='New England City and Town Area'
                        NECTADIV='New England City and Town Area Division'
                        CDCURR  ='Current Congressional District'
                        SLDU    ='State Legislative District Upper'
                        SLDL    ='State Legislative District Lower'
                        ZCTA5   ='ZIP Code Tabulation Area (5-digit)'
                        PUMA5   ='Public Use Microdata Area - 5% File'
                        PUMA1   ='Public Use Microdata Area - 1% File'                                          ;
  INPUT
    FILEID   $ 1-6                          STUSAB   $ 7-8                          SUMLEVEL  $ 9-11

                        COMPONENT $             12-13               LOGRECNO  $ 14-20               US      $ 21-21
                        REGION    $ 22-22                   DIVISION  $ 23-23                       STATECE  $ 24-25

                        STATE     $ 26-27                   COUNTY    $ 28-30                       COUSUB   $ 31-35
                        PLACE     $      36-40              TRACT  $      41-46                     BLKGRP   $ 47-47

                        CONCIT    $ 48-52                   AIANHH    $ 53-56                       AIANHHFP $ 57-61
                        AIHHTLI   $ 62-62                   AITSCE    $ 63-65                       AITS     $ 66-70

                        ANRC      $ 71-75                   CBSA      $ 76-80          CSA      $ 81-83
                        METDIV    $ 84-88                   MACC      $ 89-89                       MEMI      $ 90-90

                        NECTA     $      91-95    CNECTA   $ 96-98              NECTADIV $ 99-103
                        UA        $ 104-108                 UACP      $ 109-113                     CDCURR   $ 114-115

                        SLDU      $ 116-118                 SLDL      $ 119-121        VTD      $ 122-127
                        ZCTA3     $ 128-130                 ZCTA5     $ 131-135                     SUBMCD   $ 136-140
```

```
                    SDELM    $ 141-145              SDSEC    $ 146-150            SDUNI    $    151-155
                    UR       $ 156-156                          PCI      $ 157-157                          TAZ      $ 158-163

                    UGA      $ 164-168                          PUMA5    $ 169-173                          PUMA1    $     174-178
                    GEOID    $ 179-218             /* GEOID is 40 char in length */
                    NAME     $ 219-418
          ;

          IF sumlevel='140'; *imports data for tracts only, similar to WHERE tract IS NOT NULL ;

run;

data work.g20135&geo (keep = STUSAB CENSUS_REGION LOGRECNO GEOID_merge STATE COUNTY TRACT);
          set work.g20135&geo;
          length CENSUS_REGION $12.;
          if                  STUSAB = 'CT' OR STUSAB = 'ME' OR STUSAB = 'MA' OR STUSAB = 'NH' OR STUSAB = 'RI'
                    OR STUSAB = 'VT' OR STUSAB = 'NJ' OR STUSAB = 'NY' OR STUSAB = 'PA'
          then do;
                    CENSUS_REGION = 'Northeast'; *assign census region;
          end;
          else if   STUSAB = 'IN' OR STUSAB = 'IL' OR STUSAB = 'MI' OR STUSAB = 'OH' OR STUSAB = 'WI'
                              OR STUSAB = 'IA' OR STUSAB = 'KS' OR STUSAB = 'MN' OR STUSAB = 'MO' OR STUSAB = 'NE'
                              OR STUSAB = 'ND' OR STUSAB = 'SD'
          then do;
                    CENSUS_REGION = 'Midwest';
          end;
          else if   STUSAB = 'DE' OR STUSAB = 'DC' OR STUSAB = 'FL' OR STUSAB = 'GA' OR STUSAB = 'MD'
                              OR STUSAB = 'NC' OR STUSAB = 'SC' OR STUSAB = 'VA' OR STUSAB = 'WV' OR STUSAB = 'AL'
                              OR STUSAB = 'KY' OR STUSAB = 'MS' OR STUSAB = 'TN' OR STUSAB = 'AR' OR STUSAB = 'LA'
                              OR STUSAB = 'OK' OR STUSAB = 'TX'
          then do;
                    CENSUS_REGION = 'South';
          end;
          else if   STUSAB = 'AZ' OR STUSAB = 'CO' OR STUSAB = 'ID' OR STUSAB = 'NM' OR STUSAB = 'MT'
                              OR STUSAB = 'UT' OR STUSAB = 'NV' OR STUSAB = 'WY' OR STUSAB = 'AK' OR STUSAB = 'CA'
                              OR STUSAB = 'HI' OR STUSAB = 'OR' OR STUSAB = 'WA'
          then do;
                    CENSUS_REGION = 'West';
          end;
          else CENSUS_REGION = 'Other';
          where tract ne '';*limit to 2013ACS_5yr census tracts only;
          length GEOID_char $12.;
          GEOID_char = CATS(STATE,COUNTY,TRACT); *format GEOID_merge to match LAT LONs GEOID_merge;
          GEOID_merge = put(input(GEOID_char,12.),12.);
run;

proc sort data=sas.Acs2013_5yr_tract_lat_long;
          by GEOID_merge;
run;

proc sort data=work.g20135&geo;
          by GEOID_merge;
run;

data work.g20135&geo.coord (keep = STUSAB CENSUS_REGION LOGRECNO STATE COUNTY TRACT GEOID_merge LAT LON); *adds internal point lat lon;
          merge work.g20135&geo(in=a) sas.Acs2013_5yr_tract_lat_long;
          by GEOID_merge;
          if a;
run;

%mend;

%macro pov_ratio_calc(geo);*calculates ratios above or below 1.5 income/poverty ratio by age group by tract. *fills tracts with 0 persons in an age class with the county-level ratio;
proc means data=work.SFe_g_0056&geo noprint;*creates a sum by county of each census poverty/income variable (for the entire county);
          class county;
          output out = work.pov_ratio_county_sum_&geo
                    sum =          CountySum_B17024e1-CountySum_B17024e131
                    ;
run;

proc sort data =work.pov_ratio_county_sum_&geo;
          by county;
run;

proc sort data =work.SFe_g_0056&geo;
          by county;
run;

data work.SFe_g_filled_co_0056&geo (drop = _TYPE_ _FREQ_);
          merge work.SFe_g_0056&geo (in=a) work.pov_ratio_county_sum_&geo;
          by county;
          if a;
run;

proc means data=work.SFe_g_0056&geo noprint;*creates a sum by state of each census poverty/income variable (for the entire state);
          class state;
          output out = work.pov_ratio_state_sum_&geo
                    sum =          StateSum_B17024e1-StateSum_B17024e131
```

```
                    ;
run;

proc sort data =work.pov_ratio_state_sum_&geo;
        by state;
run;

proc sort data =work.SFe_g_filled_co_0056&geo;
        by state;
run;

data work.SFe_g_filled_st_co_0056&geo (drop = _TYPE_ _FREQ_);
        merge work.SFe_g_filled_co_0056&geo (in=a) work.pov_ratio_state_sum_&geo;
        by state;
        if a;
run;

data  work.pov_pct_&geo;
        set work.SFe_g_filled_st_co_0056&geo;
        length      filled_e2 $26 filled_e15 $26 filled_e28 $26 filled_e41 $26 filled_e54 $26 filled_e67 $26 filled_e80 $26 filled_e93 $26
                              filled_e106 $26 filled_e119 $26;
        IF B17024e2 ^= 0 then do;
        *where age group population in a tract is not equal to zero, calculate below/above poverty ratio based on income/poverty variables for the tract using tract-level data;
                filled_e2 = 'Tract Values Used';
                pctB17024e3=B17024e3/B17024e2;
                pctB17024e4=B17024e4/B17024e2;
                pctB17024e5=B17024e5/B17024e2;
                pctB17024e6=B17024e6/B17024e2;
                pctB17024e7=B17024e7/B17024e2;
                pctB17024e8=B17024e8/B17024e2;
                pctB17024e9=B17024e9/B17024e2;
                pctB17024e10=B17024e10/B17024e2;
                pctB17024e11=B17024e11/B17024e2;
                pctB17024e12=B17024e12/B17024e2;
                pctB17024e13=B17024e13/B17024e2;
                pctB17024e14=B17024e14/B17024e2;end;
        ELSE IF CountySum_B17024e2 ^= 0 then do;
        *where age group population in a tract is zero, but the county is not equal to zero, calculate below/above poverty ratio based on income/poverty variables using county-level
data;
                filled_e2 = 'Filled with County Values';
                pctB17024e3=CountySum_B17024e3/CountySum_B17024e2;
                pctB17024e4=CountySum_B17024e4/CountySum_B17024e2;
                pctB17024e5=CountySum_B17024e5/CountySum_B17024e2;
                pctB17024e6=CountySum_B17024e6/CountySum_B17024e2;
                pctB17024e7=CountySum_B17024e7/CountySum_B17024e2;
                pctB17024e8=CountySum_B17024e8/CountySum_B17024e2;
                pctB17024e9=CountySum_B17024e9/CountySum_B17024e2;
                pctB17024e10=CountySum_B17024e10/CountySum_B17024e2;
                pctB17024e11=CountySum_B17024e11/CountySum_B17024e2;
                pctB17024e12=CountySum_B17024e12/CountySum_B17024e2;
                pctB17024e13=CountySum_B17024e13/CountySum_B17024e2;
                pctB17024e14=CountySum_B17024e14/CountySum_B17024e2;end;
        ELSE IF CountySum_B17024e2 = 0 then do;
        *where age group population in a county and tract are both zero, calculate below/above poverty ratio based on income/poverty variables using state-level data for children 17
and under;
                filled_e2 = 'Filled with State Values';
                pctB17024e3=sum(StateSum_B17024e3,StateSum_B17024e16,StateSum_B17024e29)/sum(StateSum_B17024e2,StateSum_B17024e15,StateSum_B17024e28);
                pctB17024e4=sum(StateSum_B17024e4,StateSum_B17024e17,StateSum_B17024e30)/sum(StateSum_B17024e2,StateSum_B17024e15,StateSum_B17024e28);
                pctB17024e5=sum(StateSum_B17024e5,StateSum_B17024e18,StateSum_B17024e31)/sum(StateSum_B17024e2,StateSum_B17024e15,StateSum_B17024e28);
                pctB17024e6=sum(StateSum_B17024e6,StateSum_B17024e19,StateSum_B17024e32)/sum(StateSum_B17024e2,StateSum_B17024e15,StateSum_B17024e28);
                pctB17024e7=sum(StateSum_B17024e7,StateSum_B17024e20,StateSum_B17024e33)/sum(StateSum_B17024e2,StateSum_B17024e15,StateSum_B17024e28);
                pctB17024e8=sum(StateSum_B17024e8,StateSum_B17024e21,StateSum_B17024e34)/sum(StateSum_B17024e2,StateSum_B17024e15,StateSum_B17024e28);
                pctB17024e9=sum(StateSum_B17024e9,StateSum_B17024e22,StateSum_B17024e35)/sum(StateSum_B17024e2,StateSum_B17024e15,StateSum_B17024e28);

        pctB17024e10=sum(StateSum_B17024e10,StateSum_B17024e23,StateSum_B17024e36)/sum(StateSum_B17024e2,StateSum_B17024e15,StateSum_B17024e28);

        pctB17024e11=sum(StateSum_B17024e11,StateSum_B17024e24,StateSum_B17024e37)/sum(StateSum_B17024e2,StateSum_B17024e15,StateSum_B17024e28);

        pctB17024e12=sum(StateSum_B17024e12,StateSum_B17024e25,StateSum_B17024e38)/sum(StateSum_B17024e2,StateSum_B17024e15,StateSum_B17024e28);

        pctB17024e13=sum(StateSum_B17024e13,StateSum_B17024e26,StateSum_B17024e39)/sum(StateSum_B17024e2,StateSum_B17024e15,StateSum_B17024e28);

        pctB17024e14=sum(StateSum_B17024e14,StateSum_B17024e27,StateSum_B17024e40)/sum(StateSum_B17024e2,StateSum_B17024e15,StateSum_B17024e28);end;
        IF B17024e15 ^= 0 then do;
                filled_e15 = 'Tract Values Used';
                pctB17024e16=B17024e16/B17024e15;
                pctB17024e17=B17024e17/B17024e15;
                pctB17024e18=B17024e18/B17024e15;
                pctB17024e19=B17024e19/B17024e15;
                pctB17024e20=B17024e20/B17024e15;
                pctB17024e21=B17024e21/B17024e15;
                pctB17024e22=B17024e22/B17024e15;
                pctB17024e23=B17024e23/B17024e15;
                pctB17024e24=B17024e24/B17024e15;
                pctB17024e25=B17024e25/B17024e15;
                pctB17024e26=B17024e26/B17024e15;
                pctB17024e27=B17024e27/B17024e15;end;
        ELSE IF CountySum_B17024e15 ^= 0 then do;
```

```
		filled_e15 = 'Filled with County Values';
		pctB17024e16=CountySum_B17024e16/CountySum_B17024e15;
		pctB17024e17=CountySum_B17024e17/CountySum_B17024e15;
		pctB17024e18=CountySum_B17024e18/CountySum_B17024e15;
		pctB17024e19=CountySum_B17024e19/CountySum_B17024e15;
		pctB17024e20=CountySum_B17024e20/CountySum_B17024e15;
		pctB17024e21=CountySum_B17024e21/CountySum_B17024e15;
		pctB17024e22=CountySum_B17024e22/CountySum_B17024e15;
		pctB17024e23=CountySum_B17024e23/CountySum_B17024e15;
		pctB17024e24=CountySum_B17024e24/CountySum_B17024e15;
		pctB17024e25=CountySum_B17024e25/CountySum_B17024e15;
		pctB17024e26=CountySum_B17024e26/CountySum_B17024e15;
		pctB17024e27=CountySum_B17024e27/CountySum_B17024e15;end;
ELSE IF CountySum_B17024e15 = 0 then do;
		filled_e15 = 'Filled with State Values';
		pctB17024e16=sum(StateSum_B17024e3,StateSum_B17024e16,StateSum_B17024e29)/sum(StateSum_B17024e2,StateSum_B17024e15,StateSum_B17024e28);
		pctB17024e17=sum(StateSum_B17024e4,StateSum_B17024e17,StateSum_B17024e30)/sum(StateSum_B17024e2,StateSum_B17024e15,StateSum_B17024e28);
		pctB17024e18=sum(StateSum_B17024e5,StateSum_B17024e18,StateSum_B17024e31)/sum(StateSum_B17024e2,StateSum_B17024e15,StateSum_B17024e28);
		pctB17024e19=sum(StateSum_B17024e6,StateSum_B17024e19,StateSum_B17024e32)/sum(StateSum_B17024e2,StateSum_B17024e15,StateSum_B17024e28);
		pctB17024e20=sum(StateSum_B17024e7,StateSum_B17024e20,StateSum_B17024e33)/sum(StateSum_B17024e2,StateSum_B17024e15,StateSum_B17024e28);
		pctB17024e21=sum(StateSum_B17024e8,StateSum_B17024e21,StateSum_B17024e34)/sum(StateSum_B17024e2,StateSum_B17024e15,StateSum_B17024e28);
		pctB17024e22=sum(StateSum_B17024e9,StateSum_B17024e22,StateSum_B17024e35)/sum(StateSum_B17024e2,StateSum_B17024e15,StateSum_B17024e28);

pctB17024e23=sum(StateSum_B17024e10,StateSum_B17024e23,StateSum_B17024e36)/sum(StateSum_B17024e2,StateSum_B17024e15,StateSum_B17024e28);

pctB17024e24=sum(StateSum_B17024e11,StateSum_B17024e24,StateSum_B17024e37)/sum(StateSum_B17024e2,StateSum_B17024e15,StateSum_B17024e28);

pctB17024e25=sum(StateSum_B17024e12,StateSum_B17024e25,StateSum_B17024e38)/sum(StateSum_B17024e2,StateSum_B17024e15,StateSum_B17024e28);

pctB17024e26=sum(StateSum_B17024e13,StateSum_B17024e26,StateSum_B17024e39)/sum(StateSum_B17024e2,StateSum_B17024e15,StateSum_B17024e28);

pctB17024e27=sum(StateSum_B17024e14,StateSum_B17024e27,StateSum_B17024e40)/sum(StateSum_B17024e2,StateSum_B17024e15,StateSum_B17024e28);end;
IF B17024e28 ^= 0 then do;
		filled_e28 = 'Tract Values Used';
		pctB17024e29=B17024e29/B17024e28;
		pctB17024e30=B17024e30/B17024e28;
		pctB17024e31=B17024e31/B17024e28;
		pctB17024e32=B17024e32/B17024e28;
		pctB17024e33=B17024e33/B17024e28;
		pctB17024e34=B17024e34/B17024e28;
		pctB17024e35=B17024e35/B17024e28;
		pctB17024e36=B17024e36/B17024e28;
		pctB17024e37=B17024e37/B17024e28;
		pctB17024e38=B17024e38/B17024e28;
		pctB17024e39=B17024e39/B17024e28;
		pctB17024e40=B17024e40/B17024e28;end;
ELSE IF CountySum_B17024e28 ^= 0 then do;
		filled_e28 = 'Filled with County Values';
		pctB17024e29=CountySum_B17024e29/CountySum_B17024e28;
		pctB17024e30=CountySum_B17024e30/CountySum_B17024e28;
		pctB17024e31=CountySum_B17024e31/CountySum_B17024e28;
		pctB17024e32=CountySum_B17024e32/CountySum_B17024e28;
		pctB17024e33=CountySum_B17024e33/CountySum_B17024e28;
		pctB17024e34=CountySum_B17024e34/CountySum_B17024e28;
		pctB17024e35=CountySum_B17024e35/CountySum_B17024e28;
		pctB17024e36=CountySum_B17024e36/CountySum_B17024e28;
		pctB17024e37=CountySum_B17024e37/CountySum_B17024e28;
		pctB17024e38=CountySum_B17024e38/CountySum_B17024e28;
		pctB17024e39=CountySum_B17024e39/CountySum_B17024e28;
		pctB17024e40=CountySum_B17024e40/CountySum_B17024e28;end;
ELSE IF CountySum_B17024e28 = 0 then do;
		filled_e28 = 'Filled with State Values';
		pctB17024e29=sum(StateSum_B17024e3,StateSum_B17024e16,StateSum_B17024e29)/sum(StateSum_B17024e2,StateSum_B17024e15,StateSum_B17024e28);
		pctB17024e30=sum(StateSum_B17024e4,StateSum_B17024e17,StateSum_B17024e30)/sum(StateSum_B17024e2,StateSum_B17024e15,StateSum_B17024e28);
		pctB17024e31=sum(StateSum_B17024e5,StateSum_B17024e18,StateSum_B17024e31)/sum(StateSum_B17024e2,StateSum_B17024e15,StateSum_B17024e28);
		pctB17024e32=sum(StateSum_B17024e6,StateSum_B17024e19,StateSum_B17024e32)/sum(StateSum_B17024e2,StateSum_B17024e15,StateSum_B17024e28);
		pctB17024e33=sum(StateSum_B17024e7,StateSum_B17024e20,StateSum_B17024e33)/sum(StateSum_B17024e2,StateSum_B17024e15,StateSum_B17024e28);
		pctB17024e34=sum(StateSum_B17024e8,StateSum_B17024e21,StateSum_B17024e34)/sum(StateSum_B17024e2,StateSum_B17024e15,StateSum_B17024e28);
		pctB17024e35=sum(StateSum_B17024e9,StateSum_B17024e22,StateSum_B17024e35)/sum(StateSum_B17024e2,StateSum_B17024e15,StateSum_B17024e28);

pctB17024e36=sum(StateSum_B17024e10,StateSum_B17024e23,StateSum_B17024e36)/sum(StateSum_B17024e2,StateSum_B17024e15,StateSum_B17024e28);

pctB17024e37=sum(StateSum_B17024e11,StateSum_B17024e24,StateSum_B17024e37)/sum(StateSum_B17024e2,StateSum_B17024e15,StateSum_B17024e28);

pctB17024e38=sum(StateSum_B17024e12,StateSum_B17024e25,StateSum_B17024e38)/sum(StateSum_B17024e2,StateSum_B17024e15,StateSum_B17024e28);

pctB17024e39=sum(StateSum_B17024e13,StateSum_B17024e26,StateSum_B17024e39)/sum(StateSum_B17024e2,StateSum_B17024e15,StateSum_B17024e28);

pctB17024e40=sum(StateSum_B17024e14,StateSum_B17024e27,StateSum_B17024e40)/sum(StateSum_B17024e2,StateSum_B17024e15,StateSum_B17024e28);end;
IF B17024e41 ^= 0 then do;
		filled_e41 = 'Tract Values Used';
		pctB17024e42=B17024e42/B17024e41;
		pctB17024e43=B17024e43/B17024e41;
		pctB17024e44=B17024e44/B17024e41;
		pctB17024e45=B17024e45/B17024e41;
		pctB17024e46=B17024e46/B17024e41;
		pctB17024e47=B17024e47/B17024e41;
		pctB17024e48=B17024e48/B17024e41;
```

```
			pctB17024e49=B17024e49/B17024e41;
			pctB17024e50=B17024e50/B17024e41;
			pctB17024e51=B17024e51/B17024e41;
			pctB17024e52=B17024e52/B17024e41;
			pctB17024e53=B17024e53/B17024e41;end;
		ELSE IF CountySum_B17024e41 ^= 0 then do;
			filled_e41 = 'Filled with County Values';
			pctB17024e42=CountySum_B17024e42/CountySum_B17024e41;
			pctB17024e43=CountySum_B17024e43/CountySum_B17024e41;
			pctB17024e44=CountySum_B17024e44/CountySum_B17024e41;
			pctB17024e45=CountySum_B17024e45/CountySum_B17024e41;
			pctB17024e46=CountySum_B17024e46/CountySum_B17024e41;
			pctB17024e47=CountySum_B17024e47/CountySum_B17024e41;
			pctB17024e48=CountySum_B17024e48/CountySum_B17024e41;
			pctB17024e49=CountySum_B17024e49/CountySum_B17024e41;
			pctB17024e50=CountySum_B17024e50/CountySum_B17024e41;
			pctB17024e51=CountySum_B17024e51/CountySum_B17024e41;
			pctB17024e52=CountySum_B17024e52/CountySum_B17024e41;
			pctB17024e53=CountySum_B17024e53/CountySum_B17024e41;end;
		ELSE IF CountySum_B17024e41 = 0 then do;
		*where age group population in a county and tract are both zero, calculate below/above poverty ratio based on income/poverty variables using state-level data for adults 18
and over;
			filled_e41 = 'Filled with State Values';

		pctB17024e42=sum(StateSum_B17024e42,StateSum_B17024e55,StateSum_B17024e68,StateSum_B17024e81,StateSum_B17024e94,StateSum_B17024e107,StateSum_B17024
e120)/sum(StateSum_B17024e41,StateSum_B17024e54,StateSum_B17024e67,StateSum_B17024e80,StateSum_B17024e93,StateSum_B17024e106,StateSum_B17024e119);

		pctB17024e43=sum(StateSum_B17024e43,StateSum_B17024e56,StateSum_B17024e69,StateSum_B17024e82,StateSum_B17024e95,StateSum_B17024e108,StateSum_B17024
e121)/sum(StateSum_B17024e41,StateSum_B17024e54,StateSum_B17024e67,StateSum_B17024e80,StateSum_B17024e93,StateSum_B17024e106,StateSum_B17024e119);

		pctB17024e44=sum(StateSum_B17024e44,StateSum_B17024e57,StateSum_B17024e70,StateSum_B17024e83,StateSum_B17024e96,StateSum_B17024e109,StateSum_B17024
e122)/sum(StateSum_B17024e41,StateSum_B17024e54,StateSum_B17024e67,StateSum_B17024e80,StateSum_B17024e93,StateSum_B17024e106,StateSum_B17024e119);

		pctB17024e45=sum(StateSum_B17024e45,StateSum_B17024e58,StateSum_B17024e71,StateSum_B17024e84,StateSum_B17024e97,StateSum_B17024e110,StateSum_B17024
e123)/sum(StateSum_B17024e41,StateSum_B17024e54,StateSum_B17024e67,StateSum_B17024e80,StateSum_B17024e93,StateSum_B17024e106,StateSum_B17024e119);

		pctB17024e46=sum(StateSum_B17024e46,StateSum_B17024e59,StateSum_B17024e72,StateSum_B17024e85,StateSum_B17024e98,StateSum_B17024e111,StateSum_B17024
e124)/sum(StateSum_B17024e41,StateSum_B17024e54,StateSum_B17024e67,StateSum_B17024e80,StateSum_B17024e93,StateSum_B17024e106,StateSum_B17024e119);

		pctB17024e47=sum(StateSum_B17024e47,StateSum_B17024e60,StateSum_B17024e73,StateSum_B17024e86,StateSum_B17024e99,StateSum_B17024e112,StateSum_B17024
e125)/sum(StateSum_B17024e41,StateSum_B17024e54,StateSum_B17024e67,StateSum_B17024e80,StateSum_B17024e93,StateSum_B17024e106,StateSum_B17024e119);

		pctB17024e48=sum(StateSum_B17024e48,StateSum_B17024e61,StateSum_B17024e74,StateSum_B17024e87,StateSum_B17024e100,StateSum_B17024e113,StateSum_B1702
4e126)/sum(StateSum_B17024e41,StateSum_B17024e54,StateSum_B17024e67,StateSum_B17024e80,StateSum_B17024e93,StateSum_B17024e106,StateSum_B17024e119);

		pctB17024e49=sum(StateSum_B17024e49,StateSum_B17024e62,StateSum_B17024e75,StateSum_B17024e88,StateSum_B17024e101,StateSum_B17024e114,StateSum_B1702
4e127)/sum(StateSum_B17024e41,StateSum_B17024e54,StateSum_B17024e67,StateSum_B17024e80,StateSum_B17024e93,StateSum_B17024e106,StateSum_B17024e119);

		pctB17024e50=sum(StateSum_B17024e50,StateSum_B17024e63,StateSum_B17024e76,StateSum_B17024e89,StateSum_B17024e102,StateSum_B17024e115,StateSum_B1702
4e128)/sum(StateSum_B17024e41,StateSum_B17024e54,StateSum_B17024e67,StateSum_B17024e80,StateSum_B17024e93,StateSum_B17024e106,StateSum_B17024e119);

		pctB17024e51=sum(StateSum_B17024e51,StateSum_B17024e64,StateSum_B17024e77,StateSum_B17024e90,StateSum_B17024e103,StateSum_B17024e116,StateSum_B1702
4e129)/sum(StateSum_B17024e41,StateSum_B17024e54,StateSum_B17024e67,StateSum_B17024e80,StateSum_B17024e93,StateSum_B17024e106,StateSum_B17024e119);

		pctB17024e52=sum(StateSum_B17024e52,StateSum_B17024e65,StateSum_B17024e78,StateSum_B17024e91,StateSum_B17024e104,StateSum_B17024e117,StateSum_B1702
4e130)/sum(StateSum_B17024e41,StateSum_B17024e54,StateSum_B17024e67,StateSum_B17024e80,StateSum_B17024e93,StateSum_B17024e106,StateSum_B17024e119);

		pctB17024e53=sum(StateSum_B17024e53,StateSum_B17024e66,StateSum_B17024e79,StateSum_B17024e92,StateSum_B17024e105,StateSum_B17024e118,StateSum_B1702
4e131)/sum(StateSum_B17024e41,StateSum_B17024e54,StateSum_B17024e67,StateSum_B17024e80,StateSum_B17024e93,StateSum_B17024e106,StateSum_B17024e119);end;
		IF B17024e54 ^= 0 then do;
			filled_e54 = 'Tract Values Used';
			pctB17024e55=B17024e55/B17024e54;
			pctB17024e56=B17024e56/B17024e54;
			pctB17024e57=B17024e57/B17024e54;
			pctB17024e58=B17024e58/B17024e54;
			pctB17024e59=B17024e59/B17024e54;
			pctB17024e60=B17024e60/B17024e54;
			pctB17024e61=B17024e61/B17024e54;
			pctB17024e62=B17024e62/B17024e54;
			pctB17024e63=B17024e63/B17024e54;
			pctB17024e64=B17024e64/B17024e54;
			pctB17024e65=B17024e65/B17024e54;
			pctB17024e66=B17024e66/B17024e54;end;
		ELSE IF CountySum_B17024e54 ^= 0 then do;
			filled_e54 = 'Filled with County Values';
			pctB17024e55=CountySum_B17024e55/CountySum_B17024e54;
			pctB17024e56=CountySum_B17024e56/CountySum_B17024e54;
			pctB17024e57=CountySum_B17024e57/CountySum_B17024e54;
			pctB17024e58=CountySum_B17024e58/CountySum_B17024e54;
			pctB17024e59=CountySum_B17024e59/CountySum_B17024e54;
			pctB17024e60=CountySum_B17024e60/CountySum_B17024e54;
			pctB17024e61=CountySum_B17024e61/CountySum_B17024e54;
			pctB17024e62=CountySum_B17024e62/CountySum_B17024e54;
			pctB17024e63=CountySum_B17024e63/CountySum_B17024e54;
			pctB17024e64=CountySum_B17024e64/CountySum_B17024e54;
			pctB17024e65=CountySum_B17024e65/CountySum_B17024e54;
			pctB17024e66=CountySum_B17024e66/CountySum_B17024e54;end;
		ELSE IF CountySum_B17024e54 = 0 then do;
```

```
			filled_e54 = 'Filled with State Values';

		pctB17024e55=sum(StateSum_B17024e42,StateSum_B17024e55,StateSum_B17024e68,StateSum_B17024e81,StateSum_B17024e94,StateSum_B17024e107,StateSum_B17024
e120)/sum(StateSum_B17024e41,StateSum_B17024e54,StateSum_B17024e67,StateSum_B17024e80,StateSum_B17024e93,StateSum_B17024e106,StateSum_B17024e119);

		pctB17024e56=sum(StateSum_B17024e43,StateSum_B17024e56,StateSum_B17024e69,StateSum_B17024e82,StateSum_B17024e95,StateSum_B17024e108,StateSum_B17024
e121)/sum(StateSum_B17024e41,StateSum_B17024e54,StateSum_B17024e67,StateSum_B17024e80,StateSum_B17024e93,StateSum_B17024e106,StateSum_B17024e119);

		pctB17024e57=sum(StateSum_B17024e44,StateSum_B17024e57,StateSum_B17024e70,StateSum_B17024e83,StateSum_B17024e96,StateSum_B17024e109,StateSum_B17024
e122)/sum(StateSum_B17024e41,StateSum_B17024e54,StateSum_B17024e67,StateSum_B17024e80,StateSum_B17024e93,StateSum_B17024e106,StateSum_B17024e119);

		pctB17024e58=sum(StateSum_B17024e45,StateSum_B17024e58,StateSum_B17024e71,StateSum_B17024e84,StateSum_B17024e97,StateSum_B17024e110,StateSum_B17024
e123)/sum(StateSum_B17024e41,StateSum_B17024e54,StateSum_B17024e67,StateSum_B17024e80,StateSum_B17024e93,StateSum_B17024e106,StateSum_B17024e119);

		pctB17024e59=sum(StateSum_B17024e46,StateSum_B17024e59,StateSum_B17024e72,StateSum_B17024e85,StateSum_B17024e98,StateSum_B17024e111,StateSum_B17024
e124)/sum(StateSum_B17024e41,StateSum_B17024e54,StateSum_B17024e67,StateSum_B17024e80,StateSum_B17024e93,StateSum_B17024e106,StateSum_B17024e119);

		pctB17024e60=sum(StateSum_B17024e47,StateSum_B17024e60,StateSum_B17024e73,StateSum_B17024e86,StateSum_B17024e99,StateSum_B17024e112,StateSum_B17024
e125)/sum(StateSum_B17024e41,StateSum_B17024e54,StateSum_B17024e67,StateSum_B17024e80,StateSum_B17024e93,StateSum_B17024e106,StateSum_B17024e119);

		pctB17024e61=sum(StateSum_B17024e48,StateSum_B17024e61,StateSum_B17024e74,StateSum_B17024e87,StateSum_B17024e100,StateSum_B17024e113,StateSum_B1702
4e126)/sum(StateSum_B17024e41,StateSum_B17024e54,StateSum_B17024e67,StateSum_B17024e80,StateSum_B17024e93,StateSum_B17024e106,StateSum_B17024e119);

		pctB17024e62=sum(StateSum_B17024e49,StateSum_B17024e62,StateSum_B17024e75,StateSum_B17024e88,StateSum_B17024e101,StateSum_B17024e114,StateSum_B1702
4e127)/sum(StateSum_B17024e41,StateSum_B17024e54,StateSum_B17024e67,StateSum_B17024e80,StateSum_B17024e93,StateSum_B17024e106,StateSum_B17024e119);

		pctB17024e63=sum(StateSum_B17024e50,StateSum_B17024e63,StateSum_B17024e76,StateSum_B17024e89,StateSum_B17024e102,StateSum_B17024e115,StateSum_B1702
4e128)/sum(StateSum_B17024e41,StateSum_B17024e54,StateSum_B17024e67,StateSum_B17024e80,StateSum_B17024e93,StateSum_B17024e106,StateSum_B17024e119);

		pctB17024e64=sum(StateSum_B17024e51,StateSum_B17024e64,StateSum_B17024e77,StateSum_B17024e90,StateSum_B17024e103,StateSum_B17024e116,StateSum_B1702
4e129)/sum(StateSum_B17024e41,StateSum_B17024e54,StateSum_B17024e67,StateSum_B17024e80,StateSum_B17024e93,StateSum_B17024e106,StateSum_B17024e119);

		pctB17024e65=sum(StateSum_B17024e52,StateSum_B17024e65,StateSum_B17024e78,StateSum_B17024e91,StateSum_B17024e104,StateSum_B17024e117,StateSum_B1702
4e130)/sum(StateSum_B17024e41,StateSum_B17024e54,StateSum_B17024e67,StateSum_B17024e80,StateSum_B17024e93,StateSum_B17024e106,StateSum_B17024e119);

		pctB17024e66=sum(StateSum_B17024e53,StateSum_B17024e66,StateSum_B17024e79,StateSum_B17024e92,StateSum_B17024e105,StateSum_B17024e118,StateSum_B1702
4e131)/sum(StateSum_B17024e41,StateSum_B17024e54,StateSum_B17024e67,StateSum_B17024e80,StateSum_B17024e93,StateSum_B17024e106,StateSum_B17024e119);end;
		IF B17024e67 ^= 0 then do;
			filled_e67 = 'Tract Values Used';
			pctB17024e68=B17024e68/B17024e67;
			pctB17024e69=B17024e69/B17024e67;
			pctB17024e70=B17024e70/B17024e67;
			pctB17024e71=B17024e71/B17024e67;
			pctB17024e72=B17024e72/B17024e67;
			pctB17024e73=B17024e73/B17024e67;
			pctB17024e74=B17024e74/B17024e67;
			pctB17024e75=B17024e75/B17024e67;
			pctB17024e76=B17024e76/B17024e67;
			pctB17024e77=B17024e77/B17024e67;
			pctB17024e78=B17024e78/B17024e67;
			pctB17024e79=B17024e79/B17024e67;end;
		ELSE IF CountySum_B17024e67 ^= 0 then do;
			filled_e67 = 'Filled with County Values';
			pctB17024e68=CountySum_B17024e68/CountySum_B17024e67;
			pctB17024e69=CountySum_B17024e69/CountySum_B17024e67;
			pctB17024e70=CountySum_B17024e70/CountySum_B17024e67;
			pctB17024e71=CountySum_B17024e71/CountySum_B17024e67;
			pctB17024e72=CountySum_B17024e72/CountySum_B17024e67;
			pctB17024e73=CountySum_B17024e73/CountySum_B17024e67;
			pctB17024e74=CountySum_B17024e74/CountySum_B17024e67;
			pctB17024e75=CountySum_B17024e75/CountySum_B17024e67;
			pctB17024e76=CountySum_B17024e76/CountySum_B17024e67;
			pctB17024e77=CountySum_B17024e77/CountySum_B17024e67;
			pctB17024e78=CountySum_B17024e78/CountySum_B17024e67;
			pctB17024e79=CountySum_B17024e79/CountySum_B17024e67;end;
		ELSE IF CountySum_B17024e67 = 0 then do;
			filled_e67 = 'Filled with State Values';

		pctB17024e68=sum(StateSum_B17024e42,StateSum_B17024e55,StateSum_B17024e68,StateSum_B17024e81,StateSum_B17024e94,StateSum_B17024e107,StateSum_B17024
e120)/sum(StateSum_B17024e41,StateSum_B17024e54,StateSum_B17024e67,StateSum_B17024e80,StateSum_B17024e93,StateSum_B17024e106,StateSum_B17024e119);

		pctB17024e69=sum(StateSum_B17024e43,StateSum_B17024e56,StateSum_B17024e69,StateSum_B17024e82,StateSum_B17024e95,StateSum_B17024e108,StateSum_B17024
e121)/sum(StateSum_B17024e41,StateSum_B17024e54,StateSum_B17024e67,StateSum_B17024e80,StateSum_B17024e93,StateSum_B17024e106,StateSum_B17024e119);

		pctB17024e70=sum(StateSum_B17024e44,StateSum_B17024e57,StateSum_B17024e70,StateSum_B17024e83,StateSum_B17024e96,StateSum_B17024e109,StateSum_B17024
e122)/sum(StateSum_B17024e41,StateSum_B17024e54,StateSum_B17024e67,StateSum_B17024e80,StateSum_B17024e93,StateSum_B17024e106,StateSum_B17024e119);

		pctB17024e71=sum(StateSum_B17024e45,StateSum_B17024e58,StateSum_B17024e71,StateSum_B17024e84,StateSum_B17024e97,StateSum_B17024e110,StateSum_B17024
e123)/sum(StateSum_B17024e41,StateSum_B17024e54,StateSum_B17024e67,StateSum_B17024e80,StateSum_B17024e93,StateSum_B17024e106,StateSum_B17024e119);

		pctB17024e72=sum(StateSum_B17024e46,StateSum_B17024e59,StateSum_B17024e72,StateSum_B17024e85,StateSum_B17024e98,StateSum_B17024e111,StateSum_B17024
e124)/sum(StateSum_B17024e41,StateSum_B17024e54,StateSum_B17024e67,StateSum_B17024e80,StateSum_B17024e93,StateSum_B17024e106,StateSum_B17024e119);

		pctB17024e73=sum(StateSum_B17024e47,StateSum_B17024e60,StateSum_B17024e73,StateSum_B17024e86,StateSum_B17024e99,StateSum_B17024e112,StateSum_B17024
e125)/sum(StateSum_B17024e41,StateSum_B17024e54,StateSum_B17024e67,StateSum_B17024e80,StateSum_B17024e93,StateSum_B17024e106,StateSum_B17024e119);

		pctB17024e74=sum(StateSum_B17024e48,StateSum_B17024e61,StateSum_B17024e74,StateSum_B17024e87,StateSum_B17024e100,StateSum_B17024e113,StateSum_B1702
4e126)/sum(StateSum_B17024e41,StateSum_B17024e54,StateSum_B17024e67,StateSum_B17024e80,StateSum_B17024e93,StateSum_B17024e106,StateSum_B17024e119);
```

```
		pctB17024e75=sum(StateSum_B17024e49,StateSum_B17024e62,StateSum_B17024e75,StateSum_B17024e88,StateSum_B17024e101,StateSum_B17024e114,StateSum_B1702
4e127)/sum(StateSum_B17024e41,StateSum_B17024e54,StateSum_B17024e67,StateSum_B17024e80,StateSum_B17024e93,StateSum_B17024e106,StateSum_B17024e119);

		pctB17024e76=sum(StateSum_B17024e50,StateSum_B17024e63,StateSum_B17024e76,StateSum_B17024e89,StateSum_B17024e102,StateSum_B17024e115,StateSum_B1702
4e128)/sum(StateSum_B17024e41,StateSum_B17024e54,StateSum_B17024e67,StateSum_B17024e80,StateSum_B17024e93,StateSum_B17024e106,StateSum_B17024e119);

		pctB17024e77=sum(StateSum_B17024e51,StateSum_B17024e64,StateSum_B17024e77,StateSum_B17024e90,StateSum_B17024e103,StateSum_B17024e116,StateSum_B1702
4e129)/sum(StateSum_B17024e41,StateSum_B17024e54,StateSum_B17024e67,StateSum_B17024e80,StateSum_B17024e93,StateSum_B17024e106,StateSum_B17024e119);

		pctB17024e78=sum(StateSum_B17024e52,StateSum_B17024e65,StateSum_B17024e78,StateSum_B17024e91,StateSum_B17024e104,StateSum_B17024e117,StateSum_B1702
4e130)/sum(StateSum_B17024e41,StateSum_B17024e54,StateSum_B17024e67,StateSum_B17024e80,StateSum_B17024e93,StateSum_B17024e106,StateSum_B17024e119);

		pctB17024e79=sum(StateSum_B17024e53,StateSum_B17024e66,StateSum_B17024e79,StateSum_B17024e92,StateSum_B17024e105,StateSum_B17024e118,StateSum_B1702
4e131)/sum(StateSum_B17024e41,StateSum_B17024e54,StateSum_B17024e67,StateSum_B17024e80,StateSum_B17024e93,StateSum_B17024e106,StateSum_B17024e119);end;
	IF B17024e80 ^= 0 then do;
		filled_e80 = 'Tract Values Used';
		pctB17024e81=B17024e81/B17024e80;
		pctB17024e82=B17024e82/B17024e80;
		pctB17024e83=B17024e83/B17024e80;
		pctB17024e84=B17024e84/B17024e80;
		pctB17024e85=B17024e85/B17024e80;
		pctB17024e86=B17024e86/B17024e80;
		pctB17024e87=B17024e87/B17024e80;
		pctB17024e88=B17024e88/B17024e80;
		pctB17024e89=B17024e89/B17024e80;
		pctB17024e90=B17024e90/B17024e80;
		pctB17024e91=B17024e91/B17024e80;
		pctB17024e92=B17024e92/B17024e80;end;
	ELSE IF CountySum_B17024e80 ^= 0 then do;
		filled_e80 = 'Filled with County Values';
		pctB17024e81=CountySum_B17024e81/CountySum_B17024e80;
		pctB17024e82=CountySum_B17024e82/CountySum_B17024e80;
		pctB17024e83=CountySum_B17024e83/CountySum_B17024e80;
		pctB17024e84=CountySum_B17024e84/CountySum_B17024e80;
		pctB17024e85=CountySum_B17024e85/CountySum_B17024e80;
		pctB17024e86=CountySum_B17024e86/CountySum_B17024e80;
		pctB17024e87=CountySum_B17024e87/CountySum_B17024e80;
		pctB17024e88=CountySum_B17024e88/CountySum_B17024e80;
		pctB17024e89=CountySum_B17024e89/CountySum_B17024e80;
		pctB17024e90=CountySum_B17024e90/CountySum_B17024e80;
		pctB17024e91=CountySum_B17024e91/CountySum_B17024e80;
		pctB17024e92=CountySum_B17024e92/CountySum_B17024e80;end;
	ELSE IF CountySum_B17024e80 = 0 then do;
		filled_e80 = 'Filled with State Values';

		pctB17024e81=sum(StateSum_B17024e42,StateSum_B17024e55,StateSum_B17024e68,StateSum_B17024e81,StateSum_B17024e94,StateSum_B17024e107,StateSum_B17024
e120)/sum(StateSum_B17024e41,StateSum_B17024e54,StateSum_B17024e67,StateSum_B17024e80,StateSum_B17024e93,StateSum_B17024e106,StateSum_B17024e119);

		pctB17024e82=sum(StateSum_B17024e43,StateSum_B17024e56,StateSum_B17024e69,StateSum_B17024e82,StateSum_B17024e95,StateSum_B17024e108,StateSum_B17024
e121)/sum(StateSum_B17024e41,StateSum_B17024e54,StateSum_B17024e67,StateSum_B17024e80,StateSum_B17024e93,StateSum_B17024e106,StateSum_B17024e119);

		pctB17024e83=sum(StateSum_B17024e44,StateSum_B17024e57,StateSum_B17024e70,StateSum_B17024e83,StateSum_B17024e96,StateSum_B17024e109,StateSum_B17024
e122)/sum(StateSum_B17024e41,StateSum_B17024e54,StateSum_B17024e67,StateSum_B17024e80,StateSum_B17024e93,StateSum_B17024e106,StateSum_B17024e119);

		pctB17024e84=sum(StateSum_B17024e45,StateSum_B17024e58,StateSum_B17024e71,StateSum_B17024e84,StateSum_B17024e97,StateSum_B17024e110,StateSum_B17024
e123)/sum(StateSum_B17024e41,StateSum_B17024e54,StateSum_B17024e67,StateSum_B17024e80,StateSum_B17024e93,StateSum_B17024e106,StateSum_B17024e119);

		pctB17024e85=sum(StateSum_B17024e46,StateSum_B17024e59,StateSum_B17024e72,StateSum_B17024e85,StateSum_B17024e98,StateSum_B17024e111,StateSum_B17024
e124)/sum(StateSum_B17024e41,StateSum_B17024e54,StateSum_B17024e67,StateSum_B17024e80,StateSum_B17024e93,StateSum_B17024e106,StateSum_B17024e119);

		pctB17024e86=sum(StateSum_B17024e47,StateSum_B17024e60,StateSum_B17024e73,StateSum_B17024e86,StateSum_B17024e99,StateSum_B17024e112,StateSum_B17024
e125)/sum(StateSum_B17024e41,StateSum_B17024e54,StateSum_B17024e67,StateSum_B17024e80,StateSum_B17024e93,StateSum_B17024e106,StateSum_B17024e119);

		pctB17024e87=sum(StateSum_B17024e48,StateSum_B17024e61,StateSum_B17024e74,StateSum_B17024e87,StateSum_B17024e100,StateSum_B17024e113,StateSum_B1702
4e126)/sum(StateSum_B17024e41,StateSum_B17024e54,StateSum_B17024e67,StateSum_B17024e80,StateSum_B17024e93,StateSum_B17024e106,StateSum_B17024e119);

		pctB17024e88=sum(StateSum_B17024e49,StateSum_B17024e62,StateSum_B17024e75,StateSum_B17024e88,StateSum_B17024e101,StateSum_B17024e114,StateSum_B1702
4e127)/sum(StateSum_B17024e41,StateSum_B17024e54,StateSum_B17024e67,StateSum_B17024e80,StateSum_B17024e93,StateSum_B17024e106,StateSum_B17024e119);

		pctB17024e89=sum(StateSum_B17024e50,StateSum_B17024e63,StateSum_B17024e76,StateSum_B17024e89,StateSum_B17024e102,StateSum_B17024e115,StateSum_B1702
4e128)/sum(StateSum_B17024e41,StateSum_B17024e54,StateSum_B17024e67,StateSum_B17024e80,StateSum_B17024e93,StateSum_B17024e106,StateSum_B17024e119);

		pctB17024e90=sum(StateSum_B17024e51,StateSum_B17024e64,StateSum_B17024e77,StateSum_B17024e90,StateSum_B17024e103,StateSum_B17024e116,StateSum_B1702
4e129)/sum(StateSum_B17024e41,StateSum_B17024e54,StateSum_B17024e67,StateSum_B17024e80,StateSum_B17024e93,StateSum_B17024e106,StateSum_B17024e119);

		pctB17024e91=sum(StateSum_B17024e52,StateSum_B17024e65,StateSum_B17024e78,StateSum_B17024e91,StateSum_B17024e104,StateSum_B17024e117,StateSum_B1702
4e130)/sum(StateSum_B17024e41,StateSum_B17024e54,StateSum_B17024e67,StateSum_B17024e80,StateSum_B17024e93,StateSum_B17024e106,StateSum_B17024e119);

		pctB17024e92=sum(StateSum_B17024e53,StateSum_B17024e66,StateSum_B17024e79,StateSum_B17024e92,StateSum_B17024e105,StateSum_B17024e118,StateSum_B1702
4e131)/sum(StateSum_B17024e41,StateSum_B17024e54,StateSum_B17024e67,StateSum_B17024e80,StateSum_B17024e93,StateSum_B17024e106,StateSum_B17024e119);end;
	IF B17024e93 ^= 0 then do;
		filled_e93 = 'Tract Values Used';
		pctB17024e94=B17024e94/B17024e93;
		pctB17024e95=B17024e95/B17024e93;
		pctB17024e96=B17024e96/B17024e93;
		pctB17024e97=B17024e97/B17024e93;
		pctB17024e98=B17024e98/B17024e93;
```

```
pctB17024e99=B17024e99/B17024e93;
pctB17024e100=B17024e100/B17024e93;
pctB17024e101=B17024e101/B17024e93;
pctB17024e102=B17024e102/B17024e93;
pctB17024e103=B17024e103/B17024e93;
pctB17024e104=B17024e104/B17024e93;
pctB17024e105=B17024e105/B17024e93;end;
ELSE IF CountySum_B17024e93 ^= 0 then do;
filled_e93 = 'Filled with County Values';
pctB17024e94=CountySum_B17024e94/CountySum_B17024e93;
pctB17024e95=CountySum_B17024e95/CountySum_B17024e93;
pctB17024e96=CountySum_B17024e96/CountySum_B17024e93;
pctB17024e97=CountySum_B17024e97/CountySum_B17024e93;
pctB17024e98=CountySum_B17024e98/CountySum_B17024e93;
pctB17024e99=CountySum_B17024e99/CountySum_B17024e93;
pctB17024e100=CountySum_B17024e100/CountySum_B17024e93;
pctB17024e101=CountySum_B17024e101/CountySum_B17024e93;
pctB17024e102=CountySum_B17024e102/CountySum_B17024e93;
pctB17024e103=CountySum_B17024e103/CountySum_B17024e93;
pctB17024e104=CountySum_B17024e104/CountySum_B17024e93;
pctB17024e105=CountySum_B17024e105/CountySum_B17024e93;end;
ELSE IF CountySum_B17024e93 = 0 then do;
filled_e93 = 'Filled with State Values';

pctB17024e94=sum(StateSum_B17024e42,StateSum_B17024e55,StateSum_B17024e68,StateSum_B17024e81,StateSum_B17024e94,StateSum_B17024e107,StateSum_B17024
e120)/sum(StateSum_B17024e41,StateSum_B17024e54,StateSum_B17024e67,StateSum_B17024e80,StateSum_B17024e93,StateSum_B17024e106,StateSum_B17024e119);

pctB17024e95=sum(StateSum_B17024e43,StateSum_B17024e56,StateSum_B17024e69,StateSum_B17024e82,StateSum_B17024e95,StateSum_B17024e108,StateSum_B17024
e121)/sum(StateSum_B17024e41,StateSum_B17024e54,StateSum_B17024e67,StateSum_B17024e80,StateSum_B17024e93,StateSum_B17024e106,StateSum_B17024e119);

pctB17024e96=sum(StateSum_B17024e44,StateSum_B17024e57,StateSum_B17024e70,StateSum_B17024e83,StateSum_B17024e96,StateSum_B17024e109,StateSum_B17024
e122)/sum(StateSum_B17024e41,StateSum_B17024e54,StateSum_B17024e67,StateSum_B17024e80,StateSum_B17024e93,StateSum_B17024e106,StateSum_B17024e119);

pctB17024e97=sum(StateSum_B17024e45,StateSum_B17024e58,StateSum_B17024e71,StateSum_B17024e84,StateSum_B17024e97,StateSum_B17024e110,StateSum_B17024
e123)/sum(StateSum_B17024e41,StateSum_B17024e54,StateSum_B17024e67,StateSum_B17024e80,StateSum_B17024e93,StateSum_B17024e106,StateSum_B17024e119);

pctB17024e98=sum(StateSum_B17024e46,StateSum_B17024e59,StateSum_B17024e72,StateSum_B17024e85,StateSum_B17024e98,StateSum_B17024e111,StateSum_B17024
e124)/sum(StateSum_B17024e41,StateSum_B17024e54,StateSum_B17024e67,StateSum_B17024e80,StateSum_B17024e93,StateSum_B17024e106,StateSum_B17024e119);

pctB17024e99=sum(StateSum_B17024e47,StateSum_B17024e60,StateSum_B17024e73,StateSum_B17024e86,StateSum_B17024e99,StateSum_B17024e112,StateSum_B17024
e125)/sum(StateSum_B17024e41,StateSum_B17024e54,StateSum_B17024e67,StateSum_B17024e80,StateSum_B17024e93,StateSum_B17024e106,StateSum_B17024e119);

pctB17024e100=sum(StateSum_B17024e48,StateSum_B17024e61,StateSum_B17024e74,StateSum_B17024e87,StateSum_B17024e100,StateSum_B17024e113,StateSum_B170
24e126)/sum(StateSum_B17024e41,StateSum_B17024e54,StateSum_B17024e67,StateSum_B17024e80,StateSum_B17024e93,StateSum_B17024e106,StateSum_B17024e119);

pctB17024e101=sum(StateSum_B17024e49,StateSum_B17024e62,StateSum_B17024e75,StateSum_B17024e88,StateSum_B17024e101,StateSum_B17024e114,StateSum_B170
24e127)/sum(StateSum_B17024e41,StateSum_B17024e54,StateSum_B17024e67,StateSum_B17024e80,StateSum_B17024e93,StateSum_B17024e106,StateSum_B17024e119);

pctB17024e102=sum(StateSum_B17024e50,StateSum_B17024e63,StateSum_B17024e76,StateSum_B17024e89,StateSum_B17024e102,StateSum_B17024e115,StateSum_B170
24e128)/sum(StateSum_B17024e41,StateSum_B17024e54,StateSum_B17024e67,StateSum_B17024e80,StateSum_B17024e93,StateSum_B17024e106,StateSum_B17024e119);

pctB17024e103=sum(StateSum_B17024e51,StateSum_B17024e64,StateSum_B17024e77,StateSum_B17024e90,StateSum_B17024e103,StateSum_B17024e116,StateSum_B170
24e129)/sum(StateSum_B17024e41,StateSum_B17024e54,StateSum_B17024e67,StateSum_B17024e80,StateSum_B17024e93,StateSum_B17024e106,StateSum_B17024e119);

pctB17024e104=sum(StateSum_B17024e52,StateSum_B17024e65,StateSum_B17024e78,StateSum_B17024e91,StateSum_B17024e104,StateSum_B17024e117,StateSum_B170
24e130)/sum(StateSum_B17024e41,StateSum_B17024e54,StateSum_B17024e67,StateSum_B17024e80,StateSum_B17024e93,StateSum_B17024e106,StateSum_B17024e119);

pctB17024e105=sum(StateSum_B17024e53,StateSum_B17024e66,StateSum_B17024e79,StateSum_B17024e92,StateSum_B17024e105,StateSum_B17024e118,StateSum_B170
24e131)/sum(StateSum_B17024e41,StateSum_B17024e54,StateSum_B17024e67,StateSum_B17024e80,StateSum_B17024e93,StateSum_B17024e106,StateSum_B17024e119);end;
IF B17024e106 ^= 0 then do;
filled_e106 = 'Tract Values Used';
pctB17024e107=B17024e107/B17024e106;
pctB17024e108=B17024e108/B17024e106;
pctB17024e109=B17024e109/B17024e106;
pctB17024e110=B17024e110/B17024e106;
pctB17024e111=B17024e111/B17024e106;
pctB17024e112=B17024e112/B17024e106;
pctB17024e113=B17024e113/B17024e106;
pctB17024e114=B17024e114/B17024e106;
pctB17024e115=B17024e115/B17024e106;
pctB17024e116=B17024e116/B17024e106;
pctB17024e117=B17024e117/B17024e106;
pctB17024e118=B17024e118/B17024e106;end;
ELSE IF CountySum_B17024e106 ^= 0 then do;
filled_e106 = 'Filled with County Values';
pctB17024e107=CountySum_B17024e107/CountySum_B17024e106;
pctB17024e108=CountySum_B17024e108/CountySum_B17024e106;
pctB17024e109=CountySum_B17024e109/CountySum_B17024e106;
pctB17024e110=CountySum_B17024e110/CountySum_B17024e106;
pctB17024e111=CountySum_B17024e111/CountySum_B17024e106;
pctB17024e112=CountySum_B17024e112/CountySum_B17024e106;
pctB17024e113=CountySum_B17024e113/CountySum_B17024e106;
pctB17024e114=CountySum_B17024e114/CountySum_B17024e106;
pctB17024e115=CountySum_B17024e115/CountySum_B17024e106;
pctB17024e116=CountySum_B17024e116/CountySum_B17024e106;
pctB17024e117=CountySum_B17024e117/CountySum_B17024e106;
pctB17024e118=CountySum_B17024e118/CountySum_B17024e106;end;
ELSE IF CountySum_B17024e106 = 0 then do;
```

```
					filled_e106 = 'Filled with State Values';

		pctB17024e107=sum(StateSum_B17024e42,StateSum_B17024e55,StateSum_B17024e68,StateSum_B17024e81,StateSum_B17024e94,StateSum_B17024e107,StateSum_B1702
4e120)/sum(StateSum_B17024e41,StateSum_B17024e54,StateSum_B17024e67,StateSum_B17024e80,StateSum_B17024e93,StateSum_B17024e106,StateSum_B17024e119);

		pctB17024e108=sum(StateSum_B17024e43,StateSum_B17024e56,StateSum_B17024e69,StateSum_B17024e82,StateSum_B17024e95,StateSum_B17024e108,StateSum_B1702
4e121)/sum(StateSum_B17024e41,StateSum_B17024e54,StateSum_B17024e67,StateSum_B17024e80,StateSum_B17024e93,StateSum_B17024e106,StateSum_B17024e119);

		pctB17024e109=sum(StateSum_B17024e44,StateSum_B17024e57,StateSum_B17024e70,StateSum_B17024e83,StateSum_B17024e96,StateSum_B17024e109,StateSum_B1702
4e122)/sum(StateSum_B17024e41,StateSum_B17024e54,StateSum_B17024e67,StateSum_B17024e80,StateSum_B17024e93,StateSum_B17024e106,StateSum_B17024e119);

		pctB17024e110=sum(StateSum_B17024e45,StateSum_B17024e58,StateSum_B17024e71,StateSum_B17024e84,StateSum_B17024e97,StateSum_B17024e110,StateSum_B1702
4e123)/sum(StateSum_B17024e41,StateSum_B17024e54,StateSum_B17024e67,StateSum_B17024e80,StateSum_B17024e93,StateSum_B17024e106,StateSum_B17024e119);

		pctB17024e111=sum(StateSum_B17024e46,StateSum_B17024e59,StateSum_B17024e72,StateSum_B17024e85,StateSum_B17024e98,StateSum_B17024e111,StateSum_B1702
4e124)/sum(StateSum_B17024e41,StateSum_B17024e54,StateSum_B17024e67,StateSum_B17024e80,StateSum_B17024e93,StateSum_B17024e106,StateSum_B17024e119);

		pctB17024e112=sum(StateSum_B17024e47,StateSum_B17024e60,StateSum_B17024e73,StateSum_B17024e86,StateSum_B17024e99,StateSum_B17024e112,StateSum_B1702
4e125)/sum(StateSum_B17024e41,StateSum_B17024e54,StateSum_B17024e67,StateSum_B17024e80,StateSum_B17024e93,StateSum_B17024e106,StateSum_B17024e119);

		pctB17024e113=sum(StateSum_B17024e48,StateSum_B17024e61,StateSum_B17024e74,StateSum_B17024e87,StateSum_B17024e100,StateSum_B17024e113,StateSum_B170
24e126)/sum(StateSum_B17024e41,StateSum_B17024e54,StateSum_B17024e67,StateSum_B17024e80,StateSum_B17024e93,StateSum_B17024e106,StateSum_B17024e119);

		pctB17024e114=sum(StateSum_B17024e49,StateSum_B17024e62,StateSum_B17024e75,StateSum_B17024e88,StateSum_B17024e101,StateSum_B17024e114,StateSum_B170
24e127)/sum(StateSum_B17024e41,StateSum_B17024e54,StateSum_B17024e67,StateSum_B17024e80,StateSum_B17024e93,StateSum_B17024e106,StateSum_B17024e119);

		pctB17024e115=sum(StateSum_B17024e50,StateSum_B17024e63,StateSum_B17024e76,StateSum_B17024e89,StateSum_B17024e102,StateSum_B17024e115,StateSum_B170
24e128)/sum(StateSum_B17024e41,StateSum_B17024e54,StateSum_B17024e67,StateSum_B17024e80,StateSum_B17024e93,StateSum_B17024e106,StateSum_B17024e119);

		pctB17024e116=sum(StateSum_B17024e51,StateSum_B17024e64,StateSum_B17024e77,StateSum_B17024e90,StateSum_B17024e103,StateSum_B17024e116,StateSum_B170
24e129)/sum(StateSum_B17024e41,StateSum_B17024e54,StateSum_B17024e67,StateSum_B17024e80,StateSum_B17024e93,StateSum_B17024e106,StateSum_B17024e119);

		pctB17024e117=sum(StateSum_B17024e52,StateSum_B17024e65,StateSum_B17024e78,StateSum_B17024e91,StateSum_B17024e104,StateSum_B17024e117,StateSum_B170
24e130)/sum(StateSum_B17024e41,StateSum_B17024e54,StateSum_B17024e67,StateSum_B17024e80,StateSum_B17024e93,StateSum_B17024e106,StateSum_B17024e119);

		pctB17024e118=sum(StateSum_B17024e53,StateSum_B17024e66,StateSum_B17024e79,StateSum_B17024e92,StateSum_B17024e105,StateSum_B17024e118,StateSum_B170
24e131)/sum(StateSum_B17024e41,StateSum_B17024e54,StateSum_B17024e67,StateSum_B17024e80,StateSum_B17024e93,StateSum_B17024e106,StateSum_B17024e119);end;
		IF B17024e119 ^= 0 then do;
					filled_e119 = 'Tract Values Used';
					pctB17024e120=B17024e120/B17024e119;
					pctB17024e121=B17024e121/B17024e119;
					pctB17024e122=B17024e122/B17024e119;
					pctB17024e123=B17024e123/B17024e119;
					pctB17024e124=B17024e124/B17024e119;
					pctB17024e125=B17024e125/B17024e119;
					pctB17024e126=B17024e126/B17024e119;
					pctB17024e127=B17024e127/B17024e119;
					pctB17024e128=B17024e128/B17024e119;
					pctB17024e129=B17024e129/B17024e119;
					pctB17024e130=B17024e130/B17024e119;
					pctB17024e131=B17024e131/B17024e119;end;
		ELSE IF CountySum_B17024e119 ^= 0 then do;
					filled_e119 = 'Filled with County Values';
					pctB17024e120=CountySum_B17024e120/CountySum_B17024e119;
					pctB17024e121=CountySum_B17024e121/CountySum_B17024e119;
					pctB17024e122=CountySum_B17024e122/CountySum_B17024e119;
					pctB17024e123=CountySum_B17024e123/CountySum_B17024e119;
					pctB17024e124=CountySum_B17024e124/CountySum_B17024e119;
					pctB17024e125=CountySum_B17024e125/CountySum_B17024e119;
					pctB17024e126=CountySum_B17024e126/CountySum_B17024e119;
					pctB17024e127=CountySum_B17024e127/CountySum_B17024e119;
					pctB17024e128=CountySum_B17024e128/CountySum_B17024e119;
					pctB17024e129=CountySum_B17024e129/CountySum_B17024e119;
					pctB17024e130=CountySum_B17024e130/CountySum_B17024e119;
					pctB17024e131=CountySum_B17024e131/CountySum_B17024e119;end;
		ELSE IF CountySum_B17024e119 = 0 then do;
					filled_e119 = 'Filled with State Values';

		pctB17024e120=sum(StateSum_B17024e42,StateSum_B17024e55,StateSum_B17024e68,StateSum_B17024e81,StateSum_B17024e94,StateSum_B17024e107,StateSum_B1702
4e120)/sum(StateSum_B17024e41,StateSum_B17024e54,StateSum_B17024e67,StateSum_B17024e80,StateSum_B17024e93,StateSum_B17024e106,StateSum_B17024e119);

		pctB17024e121=sum(StateSum_B17024e43,StateSum_B17024e56,StateSum_B17024e69,StateSum_B17024e82,StateSum_B17024e95,StateSum_B17024e108,StateSum_B1702
4e121)/sum(StateSum_B17024e41,StateSum_B17024e54,StateSum_B17024e67,StateSum_B17024e80,StateSum_B17024e93,StateSum_B17024e106,StateSum_B17024e119);

		pctB17024e122=sum(StateSum_B17024e44,StateSum_B17024e57,StateSum_B17024e70,StateSum_B17024e83,StateSum_B17024e96,StateSum_B17024e109,StateSum_B1702
4e122)/sum(StateSum_B17024e41,StateSum_B17024e54,StateSum_B17024e67,StateSum_B17024e80,StateSum_B17024e93,StateSum_B17024e106,StateSum_B17024e119);

		pctB17024e123=sum(StateSum_B17024e45,StateSum_B17024e58,StateSum_B17024e71,StateSum_B17024e84,StateSum_B17024e97,StateSum_B17024e110,StateSum_B1702
4e123)/sum(StateSum_B17024e41,StateSum_B17024e54,StateSum_B17024e67,StateSum_B17024e80,StateSum_B17024e93,StateSum_B17024e106,StateSum_B17024e119);

		pctB17024e124=sum(StateSum_B17024e46,StateSum_B17024e59,StateSum_B17024e72,StateSum_B17024e85,StateSum_B17024e98,StateSum_B17024e111,StateSum_B1702
4e124)/sum(StateSum_B17024e41,StateSum_B17024e54,StateSum_B17024e67,StateSum_B17024e80,StateSum_B17024e93,StateSum_B17024e106,StateSum_B17024e119);

		pctB17024e125=sum(StateSum_B17024e47,StateSum_B17024e60,StateSum_B17024e73,StateSum_B17024e86,StateSum_B17024e99,StateSum_B17024e112,StateSum_B1702
4e125)/sum(StateSum_B17024e41,StateSum_B17024e54,StateSum_B17024e67,StateSum_B17024e80,StateSum_B17024e93,StateSum_B17024e106,StateSum_B17024e119);

		pctB17024e126=sum(StateSum_B17024e48,StateSum_B17024e61,StateSum_B17024e74,StateSum_B17024e87,StateSum_B17024e100,StateSum_B17024e113,StateSum_B170
24e126)/sum(StateSum_B17024e41,StateSum_B17024e54,StateSum_B17024e67,StateSum_B17024e80,StateSum_B17024e93,StateSum_B17024e106,StateSum_B17024e119);
```

```
	pctB17024e127=sum(StateSum_B17024e49,StateSum_B17024e62,StateSum_B17024e75,StateSum_B17024e88,StateSum_B17024e101,StateSum_B17024e114,StateSum_B17024e127)/sum(StateSum_B17024e41,StateSum_B17024e54,StateSum_B17024e67,StateSum_B17024e80,StateSum_B17024e93,StateSum_B17024e106,StateSum_B17024e119);

	pctB17024e128=sum(StateSum_B17024e50,StateSum_B17024e63,StateSum_B17024e76,StateSum_B17024e89,StateSum_B17024e102,StateSum_B17024e115,StateSum_B17024e128)/sum(StateSum_B17024e41,StateSum_B17024e54,StateSum_B17024e67,StateSum_B17024e80,StateSum_B17024e93,StateSum_B17024e106,StateSum_B17024e119);

	pctB17024e129=sum(StateSum_B17024e51,StateSum_B17024e64,StateSum_B17024e77,StateSum_B17024e90,StateSum_B17024e103,StateSum_B17024e116,StateSum_B17024e129)/sum(StateSum_B17024e41,StateSum_B17024e54,StateSum_B17024e67,StateSum_B17024e80,StateSum_B17024e93,StateSum_B17024e106,StateSum_B17024e119);

	pctB17024e130=sum(StateSum_B17024e52,StateSum_B17024e65,StateSum_B17024e78,StateSum_B17024e91,StateSum_B17024e104,StateSum_B17024e117,StateSum_B17024e130)/sum(StateSum_B17024e41,StateSum_B17024e54,StateSum_B17024e67,StateSum_B17024e80,StateSum_B17024e93,StateSum_B17024e106,StateSum_B17024e119);

	pctB17024e131=sum(StateSum_B17024e53,StateSum_B17024e66,StateSum_B17024e79,StateSum_B17024e92,StateSum_B17024e105,StateSum_B17024e118,StateSum_B17024e131)/sum(StateSum_B17024e41,StateSum_B17024e54,StateSum_B17024e67,StateSum_B17024e80,StateSum_B17024e93,StateSum_B17024e106,StateSum_B17024e119);end;
run;

data work.pov_ratio_&geo	/*calculates percents at or above poverty defined as +/- 1.5 income/poverty ratio */
				(keep=	STUSAB CENSUS_REGION LOGRECNO STATE COUNTY TRACT GEOID_merge LAT LON
					/*B17024e2 B17024e15 B17024e28 B17024e41 B17024e54 B17024e67 B17024e80 B17024e93 B17024e106 B17024e119*/
					filled_e2 filled_e15 filled_e28 filled_e41 filled_e54 filled_e67 filled_e80 filled_e93 filled_e106 filled_e119
					p0-p17 np0-np17
					p5u p6to11 p12to17 p18to24 p25to34 p35to44 p45to54 p55to64 p65to74 p75plus
					np5u np6to11 np12to17 np18to24 np25to34 np35to44 np45to54 np55to64 np65to74 np75plus);
 set work.pov_pct_&geo;
	/*The first sum provides the prob < 1.5 pov ratio with 'p' meaning poverty.		The second sum is > 1.5 pov with 'np' meaning not poverty. */
	p5u=sum(pctB17024e3,pctB17024e4,pctB17024e5,pctB17024e6,pctB17024e7);
	np5u=sum(pctB17024e8,pctB17024e9,pctB17024e10,pctB17024e11,pctB17024e12,pctB17024e13,pctB17024e14);
	p6to11=sum(pctB17024e16,pctB17024e17,pctB17024e18,pctB17024e19,pctB17024e20);
	np6to11=sum(pctB17024e21,pctB17024e22,pctB17024e23,pctB17024e24,pctB17024e25,pctB17024e26,pctB17024e27);
	p12to17=sum(pctB17024e29,pctB17024e30,pctB17024e31,pctB17024e32,pctB17024e33);
	np12to17=sum(pctB17024e34,pctB17024e35,pctB17024e36,pctB17024e37,pctB17024e38,pctB17024e39,pctB17024e40);
	p18to24=sum(pctB17024e42,pctB17024e43,pctB17024e44,pctB17024e45,pctB17024e46);
	np18to24=sum(pctB17024e47,pctB17024e48,pctB17024e49,pctB17024e50,pctB17024e51,pctB17024e52,pctB17024e53);
	p25to34=sum(pctB17024e55,pctB17024e56,pctB17024e57,pctB17024e58,pctB17024e59);
	np25to34=sum(pctB17024e60,pctB17024e61,pctB17024e62,pctB17024e63,pctB17024e64,pctB17024e65,pctB17024e66);
	p35to44=sum(pctB17024e68,pctB17024e69,pctB17024e70,pctB17024e71,pctB17024e72);
	np35to44=sum(pctB17024e73,pctB17024e74,pctB17024e75,pctB17024e76,pctB17024e77,pctB17024e78,pctB17024e79);
	p45to54=sum(pctB17024e81,pctB17024e82,pctB17024e83,pctB17024e84,pctB17024e85);
	np45to54=sum(pctB17024e86,pctB17024e87,pctB17024e88,pctB17024e89,pctB17024e90,pctB17024e91,pctB17024e92);
	p55to64=sum(pctB17024e94,pctB17024e95,pctB17024e96,pctB17024e97,pctB17024e98);
	np55to64=sum(pctB17024e99,pctB17024e100,pctB17024e101,pctB17024e102,pctB17024e103,pctB17024e104,pctB17024e105);
	p65to74=sum(pctB17024e107,pctB17024e108,pctB17024e109,pctB17024e110,pctB17024e111);
np65to74=sum(pctB17024e112,pctB17024e113,pctB17024e114,pctB17024e115,pctB17024e116,pctB17024e117,pctB17024e118);
	p75plus=sum(pctB17024e120,pctB17024e121,pctB17024e122,pctB17024e123,pctB17024e124);
np75plus=sum(pctB17024e125,pctB17024e126,pctB17024e127,pctB17024e128,pctB17024e129,pctB17024e130,pctB17024e131);

	/*copy the percents +/- 1.5 income/poverty ratio for ages 5 and under, 6to11, and 12to17 to separate ages 1-17 for which asthma
	prevalence data are available*/
	p0=p5u; p1=p5u; p2=p5u; p3=p5u; p4=p5u; p5=p5u;
	np0=np5u; np1=np5u; np1=np5u; np2=np5u; np3=np5u; np4=np5u; np5=np5u;
	p6=p6to11; p7=p6to11; p8=p6to11; p9=p6to11; p10=p6to11; p11=p6to11;
	np6=p6to11; np7=p6to11; np8=p6to11; np9=np6to11; np10=np6to11; np11=np6to11;
	p12=p12to17; p13=p12to17; p14=p12to17; p15=p12to17; p16=p12to17; p17=p12to17;
	np12=np12to17; np13=np12to17; np14=np12to17; np15=np12to17; np16=np12to17; np17=np12to17;

run;

data work.QA_pov_ratio_&geo /* checks that all calculated percents sum to 1 where they exist*/
				(keep=	STUSAB CENSUS_REGION LOGRECNO STATE COUNTY TRACT GEOID_merge LAT LON
					filled_e2 filled_e15 filled_e28 filled_e41 filled_e54 filled_e67 filled_e80 filled_e93 filled_e106 filled_e119
					/*B17024e2 B17024e15 B17024e28 B17024e41 B17024e54 B17024e67 B17024e80 B17024e93 B17024e106 B17024e119*/
					sum5u sum6to11 sum12to17 sum18to24 sum25to34 sum35to44 sum45to54 sum55to64 sum65to74 sum75plus);
	set work.pov_ratio_&geo;
		sum5u=p5u+np5u;
		sum6to11=p6to11+np6to11;
		sum12to17=p12to17+np12to17;
		sum18to24=p18to24+np18to24;
		sum25to34=p25to34+np25to34;
		sum35to44=p35to44+np35to44;
		sum45to54=p45to54+np45to54;
		sum55to64=p55to64+np55to64;
		sum65to74=p65to74+np65to74;
		sum75plus=p75plus+np75plus;
run;

data work.pov_ratio_&geo; *changes order of columns (variables);
 retain				STUSAB CENSUS_REGION LOGRECNO STATE COUNTY TRACT GEOID_merge LAT LON
					p0-p17
					p5u p6to11 p12to17 p18to24 p25to34 p35to44 p45to54 p55to64 p65to74 p75plus
					np0-np17
					np5u np6to11 np12to17 np18to24 np25to34 np35to44 np45to54 np55to64 np65to74 np75plus
```

```
                                                                                                    filled_e2 filled_e15 filled_e28 filled_e41 filled_e54 filled_e67 filled_e80 filled_e93 filled_e106
filled_e119;
 set work.pov_ratio_&geo;
run;
%mend;

%macro Import_Pov_Calc_Ratio(geo); *Runs macros that imports and merges income/poverty data with geographic data (by state) then calculates ratios above or below 1.5 income/poverty
ratio (by age group);
			%AnyGeo(&geo);
			%Read_poverty(&geo);

			proc sort data=work.SFe0056&geo; *sort estimate data;
					by logrecno;
			run;

			proc sort data=work.g20135&geo.coord; *sort geo data;
					by logrecno;
			run;

			data work.SFe_g_0056&geo ; *merges estimate and geo data;
			  merge  work.SFe0056&geo(in=a) work.g20135&geo.coord;
					by logrecno;
			  retain STUSAB STATE COUNTY TRACT LAT LON;
			  if a;
			run;

			%pov_ratio_calc(&geo);

			proc append base=sas.pov_acs2013_5yr data=work.pov_ratio_&geo; run;
			proc append base=sas.QA_pov_acs2013_5yr data=work.QA_pov_ratio_&geo; run;

%mend;

*runs macro for 50 United States, District of Columbia, and Puerto Rico;
%Import_Pov_Calc_Ratio(al);
%Import_Pov_Calc_Ratio(ak);
%Import_Pov_Calc_Ratio(az);
%Import_Pov_Calc_Ratio(ar);
%Import_Pov_Calc_Ratio(ca);
%Import_Pov_Calc_Ratio(co);
%Import_Pov_Calc_Ratio(ct);
%Import_Pov_Calc_Ratio(de);
%Import_Pov_Calc_Ratio(dc);
%Import_Pov_Calc_Ratio(fl);
%Import_Pov_Calc_Ratio(ga);
%Import_Pov_Calc_Ratio(hi);
%Import_Pov_Calc_Ratio(id);
%Import_Pov_Calc_Ratio(il);
%Import_Pov_Calc_Ratio(in);
%Import_Pov_Calc_Ratio(ia);
%Import_Pov_Calc_Ratio(ks);
%Import_Pov_Calc_Ratio(ky);
%Import_Pov_Calc_Ratio(la);
%Import_Pov_Calc_Ratio(me);
%Import_Pov_Calc_Ratio(md);
%Import_Pov_Calc_Ratio(ma);
%Import_Pov_Calc_Ratio(mi);
%Import_Pov_Calc_Ratio(mn);
%Import_Pov_Calc_Ratio(ms);
%Import_Pov_Calc_Ratio(mo);
%Import_Pov_Calc_Ratio(mt);
%Import_Pov_Calc_Ratio(ne);
%Import_Pov_Calc_Ratio(nv);
%Import_Pov_Calc_Ratio(nh);
%Import_Pov_Calc_Ratio(nj);
%Import_Pov_Calc_Ratio(nm);
%Import_Pov_Calc_Ratio(ny);
%Import_Pov_Calc_Ratio(nc);
%Import_Pov_Calc_Ratio(nd);
%Import_Pov_Calc_Ratio(oh);
%Import_Pov_Calc_Ratio(ok);
%Import_Pov_Calc_Ratio(or);
%Import_Pov_Calc_Ratio(pa);
%Import_Pov_Calc_Ratio(ri);
%Import_Pov_Calc_Ratio(sc);
%Import_Pov_Calc_Ratio(sd);
%Import_Pov_Calc_Ratio(tn);
%Import_Pov_Calc_Ratio(tx);
%Import_Pov_Calc_Ratio(ut);
%Import_Pov_Calc_Ratio(vt);
%Import_Pov_Calc_Ratio(va);
%Import_Pov_Calc_Ratio(wa);
%Import_Pov_Calc_Ratio(wv);
%Import_Pov_Calc_Ratio(wi);
%Import_Pov_Calc_Ratio(wy);
%Import_Pov_Calc_Ratio(pr);
```

Attachment 4 – Proc Survey Logistic Model Results - Evaluating Influence of Personal Attributes and their Interaction

Table 1. Logistic model parameter estimates, coefficient variation, and statistical significance for personal attributes that influence asthma prevalence in adults.

US REGION (degrees of freedom)	Variable	Parameter Estimate	Std Err	Statistical Significance (ProbT)
Northeast (51)	**Intercept**	-2.610	0.114	**<0.001**
	family income	0.045	0.188	0.810
	Black African American	-0.184	0.232	0.432
	BMI	0.367	0.136	**0.009**
	sex	0.418	0.091	**<0.001**
	age	-0.006	0.002	**0.002**
	family income*sex	0.345	0.236	0.150
	Black African American*sex	-0.038	0.314	0.905
	BMI*sex	0.296	0.181	0.107
	family income*BMI	0.567	0.276	**0.045**
	Black African American*BMI	-0.163	0.290	0.577
	family income*Black African American	0.675	0.499	0.183
	family income*Black African American*BMI	-0.071	0.667	0.915
	family income*BMI*sex	-0.401	0.375	0.290
	Black African American*BMI*sex	0.257	0.389	0.512
	family income*Black African American*sex	-0.298	0.648	0.647
	family income *Black African American*BMI*sex	-0.349	0.862	0.688
Midwest (66)	**Intercept**	-2.849	0.100	**<0.001**
	family income	0.543	0.137	**<0.001**
	Black African American	0.347	0.265	0.195
	BMI	0.280	0.118	**0.020**
	sex	0.420	0.072	**<0.001**
	age	-0.004	0.001	**0.012**
	family income*sex	-0.219	0.141	0.125
	Black African American*sex	0.109	0.315	0.732
	BMI*sex	0.380	0.130	**0.005**
	family income*BMI	0.122	0.237	0.609
	Black African American*BMI	-0.223	0.357	0.534
	family income*Black African American	-0.295	0.374	0.434
	family income*Black African American*BMI	0.133	0.481	0.784
	family income*BMI*sex	0.080	0.277	0.774
	Black African American*BMI*sex	-0.551	0.427	0.202
	family income*Black African American*sex	0.093	0.428	0.829
	family income *Black African American*BMI*sex	0.390	0.533	0.466

US REGION (degrees of freedom)	Variable	Parameter Estimate	Std Err	Statistical Significance (ProbT)
South (115)	**Intercept**	-3.147	0.092	**<0.001**
	family income	0.325	0.123	**0.010**
	Black African American	0.217	0.150	0.151
	BMI	0.341	0.102	**0.001**
	sex	0.540	0.076	**<0.001**
	age	-0.001	0.001	0.492
	family income*sex	-0.039	0.152	0.797
	Black African American*sex	-0.220	0.205	0.286
	BMI*sex	0.262	0.130	**0.045**
	family income*BMI	0.123	0.202	0.546
	Black African American*BMI	-0.084	0.235	0.721
	family income*Black African American	0.076	0.218	0.728
	family income*Black African American*BMI	-0.345	0.404	0.394
	family income*BMI*sex	-0.050	0.236	0.832
	Black African American*BMI*sex	0.119	0.273	0.664
	family income*Black African American*sex	0.057	0.286	0.841
	family income *Black African American*BMI*sex	0.268	0.460	0.561
West (70)	**Intercept**	-3.042	0.086	**<0.001**
	family income	0.206	0.111	0.068
	Black African American	0.432	0.295	0.147
	BMI	0.467	0.090	**<0.001**
	sex	0.532	0.069	**<0.001**
	age	0.001	0.001	0.652
	family income*sex	-0.150	0.138	0.282
	Black African American*sex	-0.812	0.314	**0.012**
	BMI*sex	0.235	0.113	**0.040**
	family income*BMI	-0.039	0.193	0.841
	Black African American*BMI	0.107	0.420	0.799
	family income*Black African American	0.164	0.441	0.711
	family income*Black African American*BMI	-0.451	0.730	0.539
	family income*BMI*sex	-0.022	0.257	0.932
	Black African American*BMI*sex	0.383	0.495	0.442
	family income*Black African American*sex	0.998	0.491	**0.046**
	family income *Black African American*BMI*sex	0.025	0.937	0.979

Table 2. Logistic model parameter estimates, coefficient variation, and statistical significance for personal attributes that influence asthma prevalence in children.

US REGION (degrees of freedom)	Variable	Parameter Estimate	Std Err	Statistical Significance (ProbT)
Northeast (51)	Intercept	-1.141	0.734	0.127
	family income	-0.378	0.298	0.210
	Black African American	0.176	0.289	0.544
	BMI	0.633	0.406	0.125
	sex	-0.314	0.202	0.127
	age	-0.063	0.047	0.188
	family income*sex	0.716	0.492	0.152
	Black African American*sex	-0.032	0.441	0.943
	BMI*sex	0.237	0.692	0.733
	family income*BMI	0.310	0.856	0.718
	Black African American*BMI	0.620	0.833	0.460
	family income*Black African American	0.300	0.459	0.517
	family income*Black African American*BMI	-0.895	1.583	0.574
	family income*BMI*sex	-0.416	1.264	0.743
	Black African American*BMI*sex	-0.647	1.320	0.626
	family income*Black African American*sex	-0.047	0.878	0.958
	family income *Black African American*BMI*sex	0.689	2.715	0.801
Midwest (66)	**Intercept**	-3.116	0.539	**<0.001**
	family income	0.626	0.275	**0.026**
	Black African American	0.648	0.308	**0.039**
	BMI	-0.047	0.338	0.889
	sex	-0.195	0.173	0.263
	age	0.052	0.032	0.114
	family income*sex	-0.313	0.352	0.378
	Black African American*sex	0.410	0.533	0.445
	BMI*sex	0.397	0.533	0.459
	family income*BMI	-0.038	0.554	0.945
	Black African American*BMI	0.740	0.827	0.374
	family income*Black African American	-0.351	0.577	0.545
	family income*Black African American*BMI	-0.679	1.303	0.604
	family income*BMI*sex	0.416	0.820	0.614
	Black African American*BMI*sex	-1.402	1.114	0.213
	family income*Black African American*sex	-0.010	0.919	0.991
	family income *Black African American*BMI*sex	1.394	1.607	0.389
South (115)	**Intercept**	-1.786	0.350	**<0.001**
	family income	0.358	0.158	**0.025**
	Black African American	1.078	0.181	**<0.001**
	BMI	0.735	0.241	**0.003**

US REGION (degrees of freedom)	Variable	Parameter Estimate	Std Err	Statistical Significance (ProbT)
	sex	0.119	0.141	0.398
	age	-0.052	0.025	**0.039**
	family income*sex	0.210	0.243	0.390
	Black African American*sex	-1.027	0.299	**0.001**
	BMI*sex	0.086	0.384	0.823
	family income*BMI	-0.740	0.434	**0.091**
	Black African American*BMI	-0.634	0.427	0.140
	family income*Black African American	-0.515	0.285	0.074
	family income*Black African American*BMI	1.153	0.637	0.073
	family income*BMI*sex	0.330	0.616	0.593
	Black African American*BMI*sex	0.767	0.806	0.343
	family income*Black African American*sex	0.346	0.446	0.439
	family income *Black African American*BMI*sex	-1.095	1.035	0.292
West (70)	**Intercept**	-1.786	0.448	**<0.001**
	family income	0.252	0.189	0.186
	Black African American	0.315	0.283	0.270
	BMI	0.123	0.335	0.714
	sex	-0.362	0.180	**0.049**
	age	-0.022	0.029	0.456
	family income*sex	-0.290	0.270	0.287
	Black African American*sex	-0.406	0.569	0.478
	BMI*sex	0.682	0.458	0.141
	family income*BMI	-0.528	0.588	0.372
	Black African American*BMI	-0.166	0.958	0.863
	family income*Black African American	0.612	0.439	0.167
	family income*Black African American*BMI	-10.865	1.411	**<0.001**
	family income*BMI*sex	-0.137	0.795	0.864
	Black African American*BMI*sex	-1.933	1.603	0.232
	family income*Black African American*sex	0.486	0.812	0.551
	family income *Black African American*BMI*sex	12.284	2.211	**<0.001**

APPENDIX F

DESCRIPTION OF THE AIR POLLUTANTS EXPOSURE MODEL (APEX)

Purpose: This Appendix briefly describes the EPA's Air Pollutants Exposure (APEX) model.

Table of Contents

F.1 Overview

APEX is the human inhalation exposure model within the Total Risk Integrated Methodology (TRIM) framework (U.S. EPA, 2017a, b). APEX is conceptually based on the probabilistic NAAQS Exposure Model (pNEM) that was used to estimate population exposures for the 1996 O_3 NAAQS review (Johnson et al., 1996a, b, c). Since that time the model has been restructured, improved, and expanded to reflect conceptual advances in the science of exposure modeling and newer input data available for the model. Key improvements to algorithms include replacement of the cohort approach with a probabilistic sampling approach focused on individuals, accounting for fatigue and oxygen debt after exercise in the calculation of ventilation rates (Isaacs et al., 2008), new approaches for construction of longitudinal activity patterns for simulated persons (Glen et al., 2008; Rosenbaum et al., 2008), and new equations for estimating resting metabolic rate (RMR) and ventilation rate (see Appendix H). Major improvements to data input to the model include updated air exchange rates (AERs), population census and commuting data, distributions of body mass and height (Appendix G), and the daily time-location-activities database (Appendix I).

APEX estimates human exposure to criteria and toxic air pollutants at local, urban, or regional scales using a stochastic, microenvironmental approach. That is, the model randomly selects data on a sample of hypothetical individuals in an actual population database and simulates each individual’s movements through time and space (e.g., at home, in vehicles) to estimate their exposure to the pollutant. APEX can assume people live and work in the same general area (i.e., that the ambient air quality is the same at home and at work) or optionally can model commuting and thus exposure at the work location for individuals who work.

The APEX model is a microenvironmental, longitudinal human exposure model for airborne pollutants. It is applied to a specified study area, which is typically a metropolitan area. The time period of the simulation is typically one year, but can easily be made either longer or shorter. APEX uses census data, such as gender and age, to generate the demographic characteristics of simulated individuals. It then assembles a composite activity diary to represent the sequence of activities and microenvironments that the individual experiences. Each microenvironment has a user-specified method for determining air quality. The inhalation exposure in each microenvironment is simply equal to the air concentration in that microenvironment. When coupled with breathing rate information and a physiological model, various measures of dose can also be calculated.

The term *microenvironment* is intended to represent the immediate surroundings of an individual, in which the pollutant of interest is assumed to be well-mixed. Time is modeled as a sequence of discrete time steps called *events*. In APEX, the concentration in a microenvironment may change between events. For each microenvironment, the user specifies the method of concentration calculation (either mass balance or regression factors, described later in this paper), the relationship of the microenvironment to the ambient air, and the strength of any pollutant sources specific to that microenvironment. Because the microenvironments that are relevant to exposure depend on the nature of the target chemical and APEX is designed to be applied to a wide range of chemicals, both the total number of microenvironments and the properties of each are free to be specified by the user.

The ambient air data are provided as input to the model in the form of time series at a list of specified locations. Typically, hourly air concentrations are used, although temporal resolutions as small as one minute may be used. The spatial range of applicability of a given ambient location is called an air district. Any number of air districts can be accommodated in a model run, subject only to computer hardware limitations. In principle, any microenvironment could be found within a given air district. Therefore, to estimate exposures as an individual engages in activities throughout the period it is necessary to determine both the microenvironment and the air district that apply for each event.

An *exposure event* is determined by the time reported in the activity diary; during any event the district, microenvironment, ambient air quality, and breathing rate are assumed to remain fixed. Since the ambient air data change every hour, the maximum duration of an event is limited to one hour. The event duration may be less than this (as short as one minute) if the activity diary indicates that the individual changes microenvironments or activities performed within the hour.

An APEX simulation includes the following steps:

(1) <u>Characterize the study area</u> - APEX selects sectors (e.g., census tracts) within a study area based on user-defined criteria and thus identifies the potentially exposed population and defines the air quality and weather input data required for the area.

(2) <u>Generate simulated individuals</u> - APEX stochastically generates a sample of simulated individuals based on the census data for the study area and human profile distribution data (such as age-specific employment probabilities). The user must specify the size of the sample. The larger the sample, the more representative it is of the population in the study area and the more stable the model results are (but also the longer the computing time).

(3) <u>Construct a long-term sequence of activity events and determine breathing rates</u> - APEX constructs an event sequence (activity pattern) spanning the period of simulation for each simulated person. The model then stochastically assigns breathing rates to each event, based on the type of activity and the physical characteristics of the simulated person.

(4) <u>Calculate pollutant concentrations in microenvironments</u> - APEX enables the user to define any microenvironment that individuals in a study area would visit. The model then calculates concentrations of each pollutant in each of the microenvironments.

(5) Calculate pollutant exposures for each simulated individual - Microenvironmental concentrations are time weighted based on individuals' events (i.e., time spent in the microenvironment) to produce a sequence of time-averaged exposures (or minute by minute time series) spanning the simulation period.

(6) Estimate dose - APEX can also calculate the dose time series for each of the simulated individuals based on the exposures and breathing rates for each event. However, dose is not needed for the SO_2 assessment and thus will not be discussed further.

(7) Estimate a health response – APEX can link an exposure-response (E-R) function generated from controlled human exposure study data with the modeled exposures to estimate the fraction of the population that could experience and adverse health outcome (e.g., lung function decrements).

The model simulation continues until exposures are determined for the user-specified number of simulated individuals. APEX then calculates population exposure statistics (such as the number of exposures exceeding user-specified levels) for the entire simulation and writes out tables of distributions of these statistics.

F.2 Model Inputs

APEX requires certain inputs from the user. The user specifies the geographic area and the range of ages and age groups to be used for the simulation. Hourly (or shorter) ambient air quality and hourly temperature data must be furnished for the entire simulation period. Other hourly meteorological data (humidity, wind speed, wind direction, precipitation) can be used by the model to estimate microenvironmental concentrations, but are optional.

In addition, most variables used in the model algorithms are represented by user-specified probability distributions which capture population variability. APEX provides great flexibility in defining model inputs and parameters, including options for the frequency of selecting new values from the probability distributions. The model also allows different distributions to be used at different times of day or on different days, and the distribution can depend conditionally on values of other parameters. The probability distributions available in APEX include beta, binary, Cauchy, discrete, exponential, extreme value, gamma, logistic, lognormal, loguniform, normal, off/on, Pareto, point (constant), triangle, uniform, Weibull, and nonparametric distributions. Minimum and maximum bounds can be specified for each distribution if a truncated distribution is appropriate. There are two options for handling truncation. The generated samples outside the truncation points can be set to the truncation limit; in this case, samples "stack up" at the truncation points. Alternatively, new random values can be selected, in which case the probability outside the limits is spread over the specified range, and thus the probabilities inside the truncation limits will be higher than the theoretical untruncated distribution.

F.3 Demographic Characteristics

The starting point for constructing a simulated individual is the population census database; this contains population counts for each combination of age, gender, race, and *sector*. The user may decide what spatial area is represented by a sector, but the default input file defines a sector as a *census tract*. Census tracts are variable in both geographic size and population number, though usually have between 1,500 and 8,000 persons. Currently, the default file contains population counts from the 2010 census for every census tract in the United States, thus the default file should be sufficient for most exposure modeling purposes. The combination of age, gender, race, and sector are selected first. The sector becomes the *home sector* for the individual, and the corresponding air district becomes the *home district*. The probabilistic selection of individuals is based on the sector population and demographic composition, and taken collectively, the set of simulated individuals constitutes a random sample from the study area.

The second step in constructing a simulated individual is to determine their employment status. This is determined by a probability which is a function of age, gender, and home sector. An input file is provided which contains employment probabilities from the 2010 census for every combination of age (16 and over), gender, and census tract. APEX assumes that persons under age 16 do not commute. For persons who are determined to be workers, APEX then randomly selects a *work sector*, based on probabilities determined from the commuting matrix. The work sector is used to assign a *work district* for the individual that may differ from the home district, and thus different ambient air quality may be used when the individual is at work.

The commuting matrix contains data on flows (number of individuals) traveling from a given home sector to a given work sector. Based on commuting data from the 2000 census, a commuting data base for the entire United States has been prepared. This permits the entire list of non-zero flows to be specified on one input file. Given a home sector, the number of destinations to which people commute varies anywhere from one to several hundred other tracts.

F.4 Attributes of Individuals

In addition to the above demographic information, each individual is assigned status and physiological attributes. The status variables are factors deemed important in estimating microenvironmental concentrations, and are specified by the user. Status variables can include, but are not limited to, people's housing type, whether their home has air conditioning, whether they use a gas stove at home, whether the stove has a gas pilot light, and whether their car has air conditioning. Physiological variables are important when estimating pollutant specific dose. These variables could include height, weight, blood volume, pulmonary diffusion rate, resting metabolic rate, energy conversion factor (liters of oxygen per kilocalorie energy expended), hemoglobin density in blood, maximum limit on metabolic equivalents of work (MET) ratios (see below), and endogenous CO production rate. All of these variables are treated probabilistically taking into account interdependencies where possible, and reflecting variability in the population.

Two key personal attributes determined for each individual in this assessment are body mass (BM) and body surface area (BSA). Each simulated individual's body mass was randomly sampled from age- and gender-specific body mass distributions generated from National Health and Nutrition Examination Survey (NHANES) data for the years 2009-2014.[1] Details in their development and the parameter values are provided in Appendix G. Then age- and gender-specific body surface area can be estimated for each simulated individual. Briefly, the BSA calculation is based on logarithmic relationships developed by Burmaster (1998) that use body mass as an independent variable as follows:

$$BSA = e^{-2.2781} BM^{0.6821} \qquad \text{Equation F-1}$$

where,

BSA = body surface area (m^2)

BM = body mass (kg)

F.5 Construction of Longitudinal Diary Sequence

The activity diary determines the sequence of microenvironments visited by the simulated person. A longitudinal sequence of daily diaries must be constructed for each simulated individual to cover the entire simulation period. The default activity diaries in APEX are derived from those in the EPA's Consolidated Human Activity Database (CHAD) (McCurdy et al., 2000; U.S. EPA 2002; 2017c), although the user could provide area specific diaries if available. There are over 55,000 CHAD diaries used for the current SO_2 assessment, each covering a 24-hour period, that have been compiled from several studies. CHAD is essentially a cross-sectional database that, for the most part, only has one diary per person. Therefore, APEX must assemble each longitudinal diary sequence for a simulated individual from many single-day diaries selected from a pool of similar people.

[1] Demographic (Demo) and Body Measurement (BMX) datasets for each of the NHANES studies were obtained from http://www.cdc.gov/nchs/nhanes/nhanes_questionnaires.htm.

APEX selects diaries from CHAD by matching gender and employment status, and by requiring that age falls within a user-specified range on either side of the age of the simulated individual. For example, if the user specifies plus or minus 20%, then for a 40-year old simulated individual, the available CHAD diaries are those from persons aged 32 to 48. Each simulated individual therefore has an age window of acceptable diaries; these windows can partially overlap those for other simulated individuals. This differs from a cohort-based approach, where the age windows are fixed and non-overlapping. The user may optionally request that APEX allow a decreased probability for selecting diaries from ages outside the primary age window, and also for selecting diaries from persons of missing gender, age, or employment status. These options allow the model to continue the simulation when diaries are not available within the primary window.

The available CHAD diaries are classified into *diary pools*, based on the temperature and day of the week. The model will select diaries from the appropriate pool for days in the simulation having matching temperature and day type characteristics. The rules for defining these pools are specified by the user. For example, the user could request that all diaries from Monday to Friday be classified together, and Saturday and Sunday diaries in another class. Alternatively, the user could instead create more than two classes of weekdays, combine all seven days into one class, or split all seven days into separate classes.

The temperature classification can be based either on daily maximum temperature, daily average temperature, or both. The user specifies both the ranges and numbers of temperatures classes. For example, the user might wish to create four temperature classes and set their ranges to below 50 °F, 50-69 °F, 70-84 °F, and above a daily maximum of 84 °F. Then day type and temperature classes are combined to create the diary pools. For example, if there are four temperature classes and two-day type classes, then there will be eight diary pools.

APEX then determines the day-type and the applicable temperature for each person's simulated day. APEX allows multiple temperature stations to be used; the sectors are automatically mapped to the nearest temperature station. This may be important for study areas such as the greater Los Angeles area, where the inland desert sectors may have very different temperatures from the coastal sectors. For selected diaries, the temperature in the home sector of the simulated person is used. For each day of the simulation, the appropriate diary pool is identified and a CHAD dairy is randomly drawn. When a diary for every day in the simulation period has been selected, they are concatenated into a single longitudinal diary covering the entire simulation for that individual. APEX contains three algorithms for stochastically selecting diaries from the pools to create the longitudinal diary. The first method selects diaries at random after stratification by age, gender, and diary pool; the second method selects diaries based on metrics related to exposure (e.g., time spent outdoors) with the goal of creating longitudinal diaries with variance properties designated by the user (Glen et al., 2008); and the third method uses a clustering algorithm to obtain more realistic recurring behavioral patterns (Rosenbaum 2008).

The final step in processing the activity diary is to map the CHAD location codes into the set of APEX microenvironments, supplied by the user as an input file. The user may define the number of microenvironments, from one up to the number of different CHAD location codes.

F.6 Key Physiological Processes Modeled

Ventilation is a general term describing the movement of air into and out of the lungs. The rate of ventilation is determined by the type of activity an individual performs which in turn is related to the amount of oxygen required to perform the activity. Minute or total ventilation rate is used to describe the volume of air moved in or out of the lungs per minute. Quantitatively, the volume of air breathed in per minute ($\dot{V}_I$) is slightly greater than the volume expired per minute ($\dot{V}_E$). Clinically, however, this difference is not important, and by convention, the ventilation rate is always measured by the expired volume.

The rate of oxygen consumption ($\dot{V}_{O2}$) is related to the rate of energy usage in performing activities as follows:

$$\dot{V}_{O2} = EE \times ECF \qquad \text{Equation (F-2)}$$

where,

$\dot{V}_{O2}$ = Oxygen consumption rate (liters O_2/minute)
EE = Energy expenditure (kcal/minute)
ECF = Energy conversion factor (liters O_2/kcal).

The ECF shows little variation and typically, commonly a value between 0.20 and 0.21 is used to represent the conversion from energy units to oxygen consumption. APEX can randomly sample from a uniform distribution defined by these lower and upper bounds to estimate an ECF for each simulated individual. The activity-specific energy expenditure is highly variable and can be estimated using metabolic equivalents (METs), or the ratios of the rate of energy consumption for non-rest activities to the resting rate of energy consumption, as follows

$$EE = MET \times RMR \qquad \text{Equation F-3}$$

where,

EE = Energy expenditure (kcal/minute)

MET = Metabolic equivalent of work (unitless)

RMR = Resting metabolic rate (kcal/minute)

APEX contains distributions of METs for all activities that might be performed by simulated individuals. APEX randomly samples from the various METs distributions to obtain values for every activity performed by each individual. Age- and sex-specific RMR are estimated once for each simulated individual using a linear regression model developed based on use BM, age, and the natural logarithms of BM and (age+1) (Equation F-4).[2] Details regarding the model derivation, ergression coefficient values, and performance evaluation are provided in Appendix H.

$$RMR = \beta_0 + \beta_1 \mathrm{BM} + \beta_2 \log(\mathrm{BM}) + \beta_3 Age + \beta_4 \log(Age) + \varepsilon_i \qquad \text{Equation F-4}$$

APEX also contains an algorithm that accounts for variability in ventilation rate ($\dot{V}_E$) due to variation in oxygen consumption ($\dot{V}_{O2}$). The approach indirectly considers influential variables such as age, sex, and body mass by use of an individual's maximum MET (or, equivalently, by VO_2m), thus the variability within age groups, and both inter- and intra-personal and variability are also accounted for. Appendix H describes this new algorithm, derived using the same clinical study data used in developing the former APEX algorithm (Graham and McCurdy, 2005), though as

$$VE = e^{(3.300 + 0.8128 \times ln\_vo2 + 0.5126 \times (VO_2 \div VO_2 m)^4 + N(0,eb) + N(0,ew))} \qquad \text{Equation F-5}$$

F.7 Estimating Microenvironmental Concentrations

The user provides rules for determining the pollutant concentration in each microenvironment. There are two available models for calculating microenvironmental concentrations: mass balance and regression factors. Any indoor microenvironment may use

[2] The "+1" modifier allows APEX to round age upwards instead of downwards to whole years, which is necessary to avoid undefined log(0) values.

either model; for each microenvironment, the user specifies whether the mass balance or factors model will be used.

F.7.1 Mass Balance Model

The mass balance method assumes that an enclosed microenvironment (e.g., a room within a home) is a single well-mixed volume in which the air concentration is approximately spatially uniform. The concentration of an air pollutant in such a microenvironment is estimated using the following four processes (and illustrated in Figure F-1):

- Inflow of air into the microenvironment;
- Outflow of air from the microenvironment;
- Removal of a pollutant from the microenvironment due to deposition, filtration, and chemical degradation; and
- Emissions from sources of a pollutant inside the microenvironment.

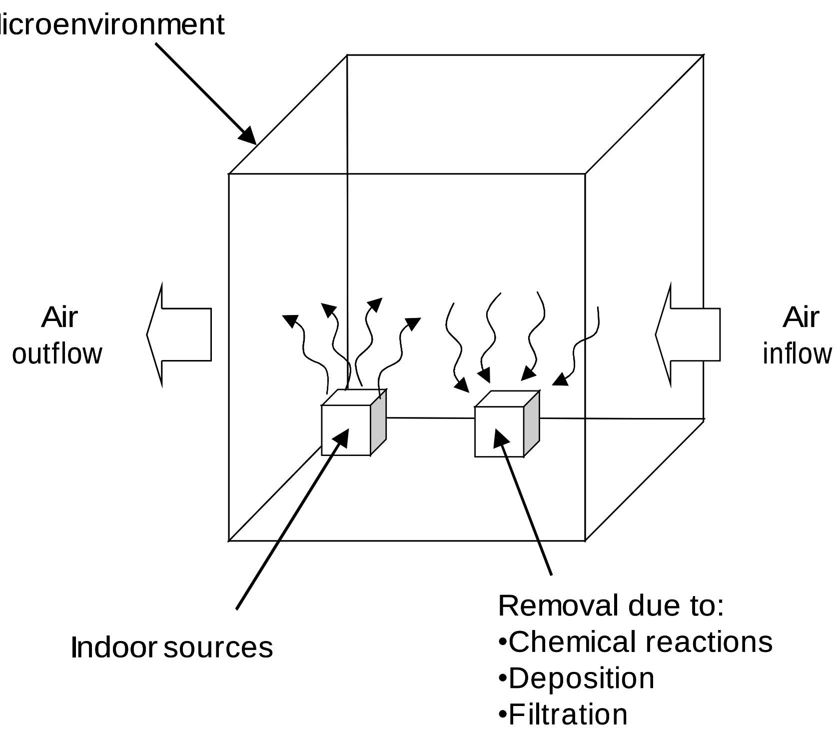


Figure F-1. Illustration of the mass balance model used by APEX.

Considering the microenvironment as a well-mixed fixed volume of air, the mass balance equation for a pollutant in the microenvironment can be written in terms of concentration:

$$\frac{dC(t)}{dt} = \dot{C}_{in} - \dot{C}_{out} - \dot{C}_{removal} + \dot{C}_{source} \qquad \text{Equation F-6}$$

where,

$C(t)$ = Concentration in the microenvironment at time t

$\dot{C}_{in}$ = Rate of change in $C(t)$ due to air entering the microenvironment

$\dot{C}_{out}$ = Rate of change in $C(t)$ due to air leaving the microenvironment

$\dot{C}_{removal}$ = Rate of change in $C(t)$ due to all internal removal processes

$\dot{C}_{source}$ = Rate of change in $C(t)$ due to all internal source terms

Concentrations are calculated in the same units as the ambient air quality data, e.g., ppm, ppb, ppt, or $\mu g/m^3$. In the following equations concentration is shown only in $\mu g/m^3$ for brevity.

The change in microenvironmental concentration due to influx of air, $\dot{C}_{in}$, is given by:

$$\dot{C}_{in} = C_{outdoor} \times f_{penetration} \times R_{air\,exchange} \qquad \text{Equation F-7}$$

where,

$C_{outdoor}$ = Ambient concentration at an outdoor microenvironment or outside an indoor microenvironment ($\mu g/m^3$)

$f_{penetration}$ = Penetration factor (unitless)

$R_{air\ exchange}$ = Air exchange rate (hr^{-1})

Because the air pressure is approximately constant in microenvironments that are modeled in practice, the flow of outside air into the microenvironment is equal to that flowing out of the microenvironment, and this flow rate is given by the air exchange rate. The air exchange rate (hr^{-1}) can be loosely interpreted as the number of times per hour the entire volume of air in the microenvironment is replaced. For some pollutants (especially particulate matter), the process of infiltration may remove a fraction of the pollutant from the outside air. The fraction that is retained in the air is given by the penetration factor $f_{penetration}$.

A proximity factor ($f_{proximity}$) and a local outdoor source term are used to account for differences in ambient concentrations between the geographic location represented by the ambient air quality data (e.g., a regional fixed-site monitor) and the geographic location of the microenvironment. That is, the outdoor air at a particular location may differ systematically from the concentration input to the model representing the air quality district. For example, a playground or house might be located next to a busy road in which case the air at the playground or outside the house would have elevated levels for mobile source pollutants such as carbon monoxide and benzene. The concentration in the air at an outdoor location or directly outside an indoor microenvironment ($C_{outdoor}$) is calculated as:

$$C_{outdoor} = f_{proximity} C_{ambient} + C_{LocalOutdoorSources} \qquad \text{Equation F-8}$$

where,

$C_{ambient}$ = Ambient air district concentration (µg/m^3)

$f_{proximity}$ = Proximity factor (unitless)

$C_{LocalOutdoorSources}$ = the contribution to the concentration at this location from local sources not represented by the ambient air district concentration (µg/m^3)

During exploratory analyses, the user may examine how a microenvironment affects overall exposure by setting the microenvironment's proximity or penetration factor to zero, thus effectively eliminating the specified microenvironment. Change in microenvironmental concentration due to outflux of air is calculated as the concentration in the microenvironment *C(t)* multiplied by the air exchange rate:

$$\dot{C}_{out} = R_{air\ exchange} \times C(t) \qquad \text{Equation F-9}$$

The third term ($\dot{C}_{removal}$) in the mass balance calculation (Equation F-6) represents removal processes within the microenvironment. There are three such processes in general: chemical reaction, deposition, and filtration. Removal can be important for pollutants such as O_3 and SO_2, for example, but not for carbon monoxide. The amount lost to chemical reactions will generally be proportional to the amount present, which in the absence of any other factors would result in an exponential decay in the concentration with time. Similarly, deposition rates are usually given by the product of a (constant) deposition velocity and a (time-varying) concentration, also resulting in an exponential decay. The third removal process is filtration, usually as part of a forced air circulation or HVAC system. Filtration will normally be more effective at removing particles than gases. In any case, filtration rates are also approximately proportional to concentration. Change in concentration due to deposition, filtration, and chemical degradation in a microenvironment is simulated based on the first-order equation:

$$\begin{aligned}\dot{C}_{removal} &= \left(R_{deposition} + R_{filtration} + R_{chemical}\right) \times C(t) \\ &= R_{removal} \times C(t)\end{aligned} \qquad \text{Equation F-10}$$

where,

$\dot{C}_{removal}$ = Change in microenvironmental concentration due to removal processes (µg/m^3/hr)
$R_{deposition}$ = Removal rate of a pollutant from a microenvironment due to deposition (hr^{-1})
$R_{filtration}$ = Removal rate of a pollutant from a microenvironment due to filtration (hr^{-1})
$R_{chemical}$ = Removal rate of a pollutant from a microenvironment due to chemical degradation (hr^{-1})
$R_{removal}$ = Removal rate of a pollutant from a microenvironment due to the combined effects of deposition, filtration, and chemical degradation (hr^{-1})

The fourth term in the mass balance calculation represents pollutant sources within the microenvironment. This is the most complicated term, in part because several sources may be present. APEX allows two methods of specifying source strengths: emission sources and concentration sources. Either may be used for mass balance microenvironments, and both can be used within the same microenvironment. The source strength values are used to calculate the term $\dot{C}_{source}$ (µg/m^3/hr).

Emission sources are expressed as emission rates in units of µg/hr, irrespective of the units of concentration. To determine the rate of change of concentration associated with an emission source S_E, it is divided by the volume of the microenvironment:

$$\dot{C}_{source,SE} = \frac{S_E}{V} \qquad \text{Equation F-11}$$

where,

$\dot{C}_{source,SE}$ = Rate of change in $C(t)$ due to the emission source S_E (µg/m^3/hr)
S_E = The emission rate (µg/hr)
V = The volume of the microenvironment (m^3)

Concentration sources (S_C) however, are expressed in units of concentration. These must be the same units as used for the ambient concentration (e.g., µg/m^3). Concentration sources are normally used as additive terms for microenvironments using the factors model. Strictly speaking, they are somewhat inconsistent with the mass balance method, since concentrations should not be inputs but should be consequences of the dynamics of the system. Nevertheless, a suitable meaning can be found by determining the rate of change of concentration ($\dot{C}_{source}$) that would result in a mean increase of S_C in the concentration, given constant parameters and equilibrium conditions, in this way:

Assume that a microenvironment is always in contact with clean air (ambient = zero), and it contains one constant concentration source. Then the mean concentration over time in this microenvironment from this source should be equal to S_C. The mean source strength expressed in ppm/hr or µg/m^3/hr is the rate of change in concentration ($\dot{C}_{source,SC}$). In equilibrium,

$$C_S = \frac{\dot{C}_{source,SC}}{R_{air\ exchange} + R_{removal}} \qquad \text{Equation F-12}$$

where, Cs is the mean increase in concentration over time in the microenvironment due to the source $\dot{C}_{source,SC}$. Thus, $\dot{C}_{source,SC}$ can be expressed as

$$\dot{C}_{source,SC} = C_S \times R_{mean} \qquad \text{Equation F-13}$$

where R_{mean} is the chemical removal rate. From Equation (F-13), R_{mean} is the sum of the air exchange rate and the removal rate ($R_{air\ exchange} + R_{removal}$) under equilibrium conditions. In general, however, the microenvironment will not be in equilibrium, but in such conditions there is no clear meaning to attach to $\dot{C}_{source,SC}$ since there is no fixed emission rate that will lead to a fixed increase in concentration. The simplest solution is to use $R_{mean} = R_{air\ exchange} + R_{removal}$. However, the user is given the option of specifically specifying R_{mean} (see discussion below). This may be used to generate a truly constant source strength $\dot{C}_{source,SC}$ by making S_C and R_{mean} both constant in time. If this is not done, then R_{mean} is simply set to the sum of ($R_{air\ exchange}$ + $R_{removal}$). If these parameters change over time, then $\dot{C}_{source,SC}$ also changes. Physically, the reason for this is that in order to maintain a fixed elevation of concentration over the base conditions, then the source emission rate would have to rise if the air exchange rate were to rise.

Multiple emission and concentration sources within a single microenvironment are combined into the final total source term by combining Equations (F-11) and (F-13):

$$\dot{C}_{source} = \dot{C}_{source,SE} + \dot{C}_{source,SC} = \frac{1}{V}\sum_{i=1}^{n_e} E_{S\,i} + R_{mean}\sum_{i=1}^{n_c} C_{S\,i} \qquad \text{Equation F-14}$$

where,

S_{Ei} = Emission source strength for emission source i (µg/hr, irrespective of the concentration units)

S_{Ci} = Emission source strength for concentration source i (µg/m^3)

n_e = Number of emission sources in the microenvironment

n_c = Number of concentration sources in the microenvironment

In Equations (F-11) and (F-14), if the units of air quality are ppm rather than µg/m^3, $1/V$ is replaced by f/V, where f = ppm / µg/m^3 = gram molecular weight / 24.45 (i.e., 24.45 being the volume (liters) of a mole of the gas at 25°C and 1 atmosphere pressure). Equations (F-7), (F-9), (F-10), and (F-14) can now be combined with Equation (F-6) to form the differential equation for the microenvironmental concentration C(t). Within the time period of a time step (at most 1 hour), $\dot{C}_{source}$ and $\dot{C}_{in}$ are assumed to be constant. Using $\dot{C}_{combined} = \dot{C}_{source} + \dot{C}_{in}$ leads to:

$$\begin{aligned}\frac{dC(t)}{dt} &= \dot{C}_{combined} - R_{air\ exchange}\, C(t) - R_{removal}\, C(t) \\ &= \dot{C}_{combined} - R_{mean}\, C(t)\end{aligned} \qquad \text{Equation F-15}$$

Solving this differential equation leads to:

$$C(t) = \frac{\dot{C}_{combined}}{R_{mean}} + \left(C(t_0) - \frac{\dot{C}_{combined}}{R_{mean}} \right) e^{-R_{mean}(t-t_0)} \qquad \text{Equation F-16}$$

where,

$C(t_0)$ = Concentration of a pollutant in a microenvironment at the beginning of a time step (µg/m^3)

$C(t)$ = Concentration of a pollutant in a microenvironment at time t within the time step (µg/m^3).

Based on Equation (F-16), the following three concentrations in a microenvironment are calculated:

$$C_{equil} = C(t \to \infty) = \frac{\dot{C}_{combined}}{R_{mean}} = \frac{\dot{C}_{source} + \dot{C}_{in}}{R_{air exchange} + R_{removal}} \qquad \text{Equation F-17}$$

$$C(t_0 + T) = C_{equil} + \left(C(t_0) - C_{equil} \right) e^{-R_{mean} T} \qquad \text{Equation F-18}$$

$$C_{mean} = \frac{1}{T}\int_{t_0}^{t_0+T} C(t)\,dt = C_{equil} + \left(C(t_0) - C_{equil}\right)\frac{1-e^{-R_{mean}T}}{R_{mean}\,T} \qquad \text{Equation F-19}$$

where,

C_{equil} = Concentration in a microenvironment (µg/m^3) if t $\rightarrow \infty$ (equilibrium state).

$C(t_0)$ = Concentration in a microenvironment at the beginning of the time step (µg/m^3)

$C_{(t0+T)}$ = Concentration in a microenvironment at the end of the time step (µg/m^3)

C_{mean} = Mean concentration over the time step in a microenvironment (µg/m^3)

R_{mean} = $R_{air\ exchange} + R_{removal}$ (hr^{-1})

At each time step of the simulation period, APEX uses Equations (F-17), (F-18), and (5A-19) to calculate the equilibrium, ending, and mean concentrations, respectively. The calculation continues to the next time step by using $C(t_0+T)$ for the previous hour as $C(t_0)$.

F.7.2 Factors Model

The factors model is simpler than the mass balance model. In this method, the value of the concentration in a microenvironment is not dependent on the concentration during the previous time step. Rather, this model uses the following equation to calculate the concentration in a microenvironment from the user-provided hourly air quality data:

$$C_{mean} = C_{ambient}\ f_{proximity}\ f_{penetration} + \sum_{i=1}^{n_c} S_{Ci} \qquad \text{Equation F-20}$$

where,

C_{mean} = Mean concentration over the time step in a microenvironment (µg/m^3)

$C_{ambient}$ = The concentration in the ambient (outdoor) environment (µg/m^3)

$f_{proximity}$ = Proximity factor (unitless)

$f_{penetration}$ = Penetration factor (unitless)

S_{Ci} = Mean air concentration resulting from source i (µg/m^3)

n_c = Number of concentration sources in the microenvironment

The user may specify distributions for proximity, penetration, and any concentration source terms. All of the parameters in Equation (F-20) are evaluated for each time step, although these values might remain constant for several time steps or even for the entire simulation.

The ambient air quality data are supplied as time series over the simulation period at several locations across the modeled region. The other variables in the factors and mass balance equations are randomly drawn from user-specified distributions. The user also controls the frequency and pattern of these random draws. Within a single day, the user selects the number of random draws to be made and the hours to which they apply. Over the simulation, the same set of 24 hourly values may either be reused on a regular basis (for example, each winter weekday), or a new set of values may be drawn. The usage patterns may depend on day of the week, on month, or both. It is also possible to define different distributions that apply if specific conditions are met. The air exchange rate is typically modeled with one set of distributions for buildings with air conditioning and another set of distributions for those which do not. The choice of a distribution within a set typically depends on the outdoor temperature and possibly other variables. In total there are eleven such *conditional variables* which can be used to select the appropriate distributions for the variables in the mass balance or factors equations.

For example, the hourly emissions of CO from a gas stove may be given by the product of three random variables: a binary on/off variable that indicates if the stove is used at all during that hour, a usage duration sampled from a continuous distribution, and an emission rate per minute of usage. The binary on/off variable may have a probability for *on* that varies by time of day and season of the year. The usage duration could be taken from a truncated normal or lognormal distribution that is resampled for each cooking event, while the emission rate could be sampled just once per stove.

F.8 Exposure and Dose Time Series Calculations

The activity diaries provide the time sequence of microenvironments visited by the simulated individual and the activities performed by each individual. The pollutant concentration in the air in each microenvironment is assumed to be spatially uniform throughout the microenvironment and unchanging within each diary event and is calculated by either the factors or the mass balance method, as specified by the user. The exposure of the individual is given by the time sequence of airborne pollutant concentrations that are encountered in the microenvironments visited. Figure F-2 illustrates the exposures for one simulated 12-year old child over a 2-day period. On both days the child travels to and from school in an automobile, goes outside to a playground in the afternoon while at school, and spends time outside at home in the evening.

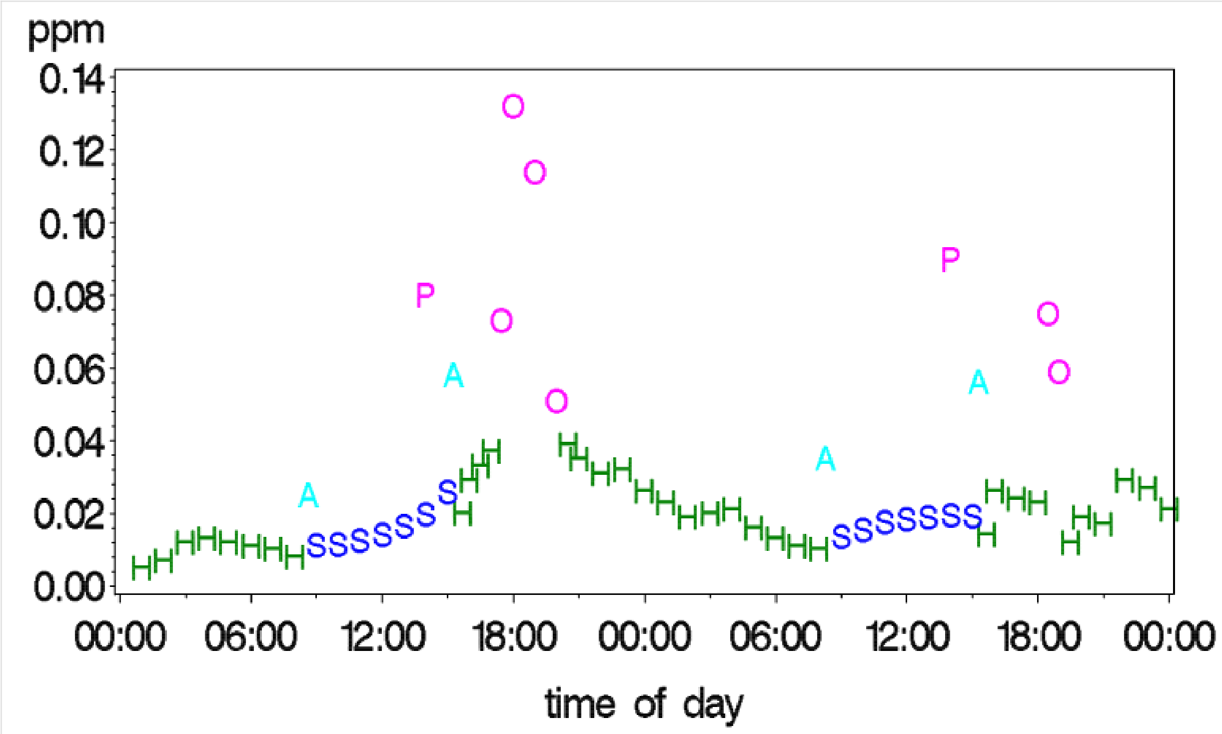


Figure F-2. Example of microenvironmental and exposure concentrations for a simulated individual over a 48-hour duration. (H: home, A: automobile, S: school, P: playground, O: outdoors at home).

In addition to exposure, APEX models breathing rates based on the physiology of each individual and the exertion levels associated with the activities performed. For each activity type in CHAD, a distribution is provided for a corresponding normalized metabolic equivalent of work or METs (McCurdy, 2000). METs are derived by dividing the metabolic energy requirements for the specific activity by a person's resting, or basal, metabolic rate. The MET ratios have less interpersonal variation than do the absolute energy expenditures. Based on age and sex, the resting metabolic rate, along with other physiological variables is determined for each individual as part of their anthropometric characteristics. Because the MET ratios are sampled independently from distributions for each diary event, it would be possible to produce time-series of MET ratios that are physiologically unrealistic. APEX employs a MET adjustment algorithm based on a modeled oxygen deficit to prevent such overestimation of MET and breathing rates (Isaacs et al., 2008). The relationship between the oxygen deficit and the applied limits on MET ratios are nonlinear and are derived from published data on work capacity and oxygen consumption. The resulting combination of microenvironmental concentration and breathing ventilation rates provides a time series of inhalation intake dose for most pollutants.

F.9 Model Output

APEX calculates the exposure and dose time series based on the events as listed on the activity diary with a minimum of one event per hour but usually more during waking hours. APEX can aggregate the event level exposure and dose time series to output hourly, daily, monthly, and annual averages. The types of output files are selected by the user, and can be as detailed as event-level data for each simulated individual (note, Figure F-3 was produced from an APEX event output file). A set of summary tables are produced for a variety of exposure and dose measures. These could include tables of person-minutes at various exposure levels, by microenvironment, a table of person-days at or above each average daily exposure level, and tables describing the distributions of exposures for different groups. An example of how APEX results can be depicted is given Figure F-3 which shows the percent of children with at least one 5-minute maximum exposure at or above different exposure levels, concomitant with moderate or greater exertion. These are results from a simulation of SO_2 exposures for Fall River, MA during 2011. From this graph it can be observed, for example, that APEX estimates 15 percent of the children in this area experienced a daily maximum 5-minute SO_2 exposure above 100 ppb while exercising, at least once during the year.

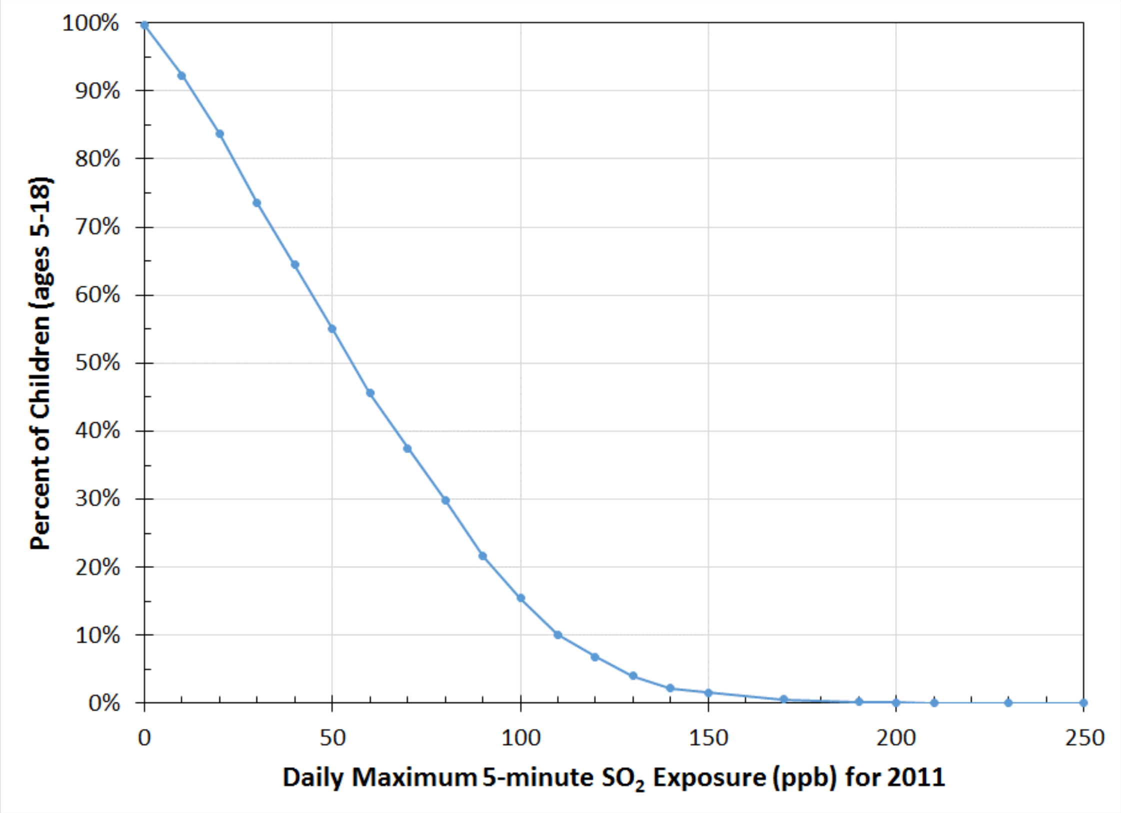


Figure F-3. The percent of simulated children (ages 5-18) experiencing at least one daily maximum 5-minute SO_2 exposure during 2011, while at moderate or greater exertion.

REFERENCES

Burmaster, D.E. (1998). LogNormal distributions for skin area as a function of body weight. *Risk Analysis*. 18(1):27-32.

Glen, G., Smith, L., Isaacs, K., McCurdy, T., Langstaff, J. (2008). A new method of longitudinal diary assembly for human exposure modeling. *J Expos Sci Environ Epidem*. 18:299-311.

Graham, S.E., McCurdy, T. (2005). Revised ventilation rate (VE) equations for use in inhalation-oriented exposure models. Report no. EPA/600/X-05/008 is Appendix A of US EPA (2009). Metabolically Derived Human Ventilation Rates: A Revised Approach Based Upon Oxygen Consumption Rates (Final Report). Report no. EPA/600/R-06/129F. Available at: http://cfpub.epa.gov/ncea/cfm/recordisplay.cfm?deid=202543.

Isaacs, K., Glen, G., McCurdy, T., Smith, L. (2008). Modeling energy expenditure and oxygen consumption in human exposure models: accounting for fatigue and EPOC. *J Expos Sci Environ Epidemiol.* 18:289-298.

Johnson, T., Capel, J., McCoy, M. (1996a). Estimation of Ozone Exposures Experienced by Urban Residents Using a Probabilistic Version of NEM and 1990 Population Data. Prepared by IT Air Quality Services for the Office of Air Quality Planning and Standards, U.S. Environmental Protection Agency, Research Triangle Park, North Carolina, September.

Johnson, T., Capel, J., Mozier, J., McCoy, M. (1996b). Estimation of Ozone Exposures Experienced by Outdoor Children in Nine Urban Areas Using a Probabilistic Version of NEM. Prepared for the Air Quality Management Division under Contract No. 68-DO-30094, April.

Johnson, T., Capel, J., McCoy, M., Mozier, J. (1996c). Estimation of Ozone Exposures Experienced by Outdoor Workers in Nine Urban Areas Using a Probabilistic Version of NEM. Prepared for the Air Quality Management Division under Contract No. 68-DO-30094, April.

McCurdy, T. (2000). Conceptual basis for multi-route intake dose modeling using an energy expenditure approach. *J Expo Anal Environ Epidemiol*.10:1-12.

McCurdy T, Glen G, Smith L, Lakkadi Y. (2000). The National Exposure Research Laboratory's Consolidated Human Activity Database. *J Expos Anal Environ Epidemiol*. 10:566-578.

Rosenbaum, A. S. (2008). The Cluster-Markov algorithm in APEX. Memorandum prepared for Stephen Graham, John Langstaff. USEPA OAQPS by ICF International.

U.S. EPA. (2002). Consolidated Human Activities Database (CHAD) Users Guide. Database and documentation available at: http://www.epa.gov/chadnet1/.

U.S. EPA. (2017a). Air Pollutants Exposure Model Documentation (APEX, Version 5) Volume I: User's Guide. Office of Air Quality Planning and Standards, U.S.

Environmental Protection Agency, Research Triangle Park, NC, 27711. EPA-452/R-17-001a. Available at: https://www.epa.gov/fera/apex-user-guides

U.S. EPA. (2017b). Air Pollutants Exposure Model Documentation (APEX, Version 5) Volume II: Technical Support Document. Office of Air Quality Planning and Standards, U.S. Environmental Protection Agency, Research Triangle Park, NC, 27711. EPA-452/R-17-001b. Available at: https://www.epa.gov/fera/apex-user-guides

U.S. EPA. (2017c). The Consolidated Human Activity Database – Master Version (CHAD-Master). Technical Memorandum. U.S. Environmental Protection Agency, National Exposure Research Laboratory, Research Triangle Park, NC, 27711. In preparation. Previous version (09/15/2014) available at: https://www.epa.gov/healthresearch/consolidated-human-activity-database-chad-use-human-exposure-and-health-studies-and

APPENDIX G

ICF FINAL MEMO: JOINT DISTRIBUTIONS OF BODY WEIGHT AND HEIGHT FOR USE IN APEX

Draft Memorandum

To:	John Langstaff, Stephen Graham, Kristin Isaacs, U.S. Environmental Protection Agency
From:	Jonathan Cohen, Graham Glen, John Hader, Chris Holder, ICF
Date:	April 20, 2017
Re:	Joint Distributions of Body Weight and Height for use in APEX (Revised from October 26, 2016 version to add Section 6).

1. Introduction and Summary

The current version of APEX uses fitted distributions for body weight (BW; also referred to as body mass) based on an analysis of the data from the National Health and Nutrition Examination Survey (NHANES) for the years 1999–2004. These distributions were developed in 2005.[1] The current version of APEX also uses fitted distributions for height (HT) based on fitted regressions for HT against age for children under 18 years of age and fitted regressions for HT against the logarithm of BW for adults 18 years and older. The regression coefficients for children depend upon the age group and gender.[2] **ICF was tasked with updating these BW fitted distributions to use more recent NHANES data and to compute parameters for the joint distribution of BW and HT**.

We downloaded and analyzed BW and HT data from NHANES for the years 2003–2014. We fitted distributions for the entire period 2003–2014 and also for the more recent period 2009–2014. As shown in Section 5, the final fitted models were very similar for the 2003–2014 and 2009–2014 periods. In this memorandum, **we present detailed results for the 2009–2014 analysis**. We provide the final parameter estimates for both groups of years in accompanying Excel spreadsheets. We can provide the detailed analyses for 2003–2014 upon request.

In Section 2, we present histograms and summary tables for the marginal distributions of BW and HT for each gender and single year of age. We compared fitted normal and log-normal distributions using the histograms and log-likelihoods and determined that **the best overall choice was a log-normal distribution for BW and a normal distribution for HT**. To allow a smooth set of parameters for different ages, **we chose the same distributional forms (but different parameters) for each combination of gender and age**.

In Section 3, **we model the joint distribution of BW and HT as a bivariate normal distribution for the HT and the logarithm of the BW, with different parameters for each age and gender**. We present scatter plots for selected single years of age.

[1] Kristin Isaacs and Luther Smith, Alion Science and Technology, “New Values for Physiological Parameters for the Exposure Model Input File Physiology.txt”. Memorandum to Tom McCurdy, EPA. December 20, 2005.

[2] Johnson T, Mihlan G, LaPointe J, Fletcher K, Capel J, Rosenbaum A, Cohen J, Stiefer P. 2000. Estimation of carbon monoxide exposures and associated carboxyhemoglobin levels for residents of Denver and Los Angeles using pNEM/CO. Appendices. EPA contract 68-D6-0064.

As shown in Section 4, the estimated parameters for each age do not vary smoothly across the ages. Therefore, **we used a natural cubic spline model to smooth each of the five parameters across the different ages for each gender**. **This approach also allowed us to smoothly extrapolate the parameters for ages 80 to 100**, since the NHANES data for recent periods combines all ages 80 and above into a single age group.

In Section 5 we compare the fitted parameters between the NHANES periods 2009–2014 and 2003–2014 and show that, after smoothing the parameters, the maximum unsigned percentage difference is 11 percent for the correlation coefficient and less than 1 percent for the means.

Finally, in Section **Error! Reference source not found.** we compare summaries of the HT, BW, and body mass index from the Personal Summary files generated by running APEX with the old and updated method for calculating height and weight. There is now a better correlation between HT, WT, and age for young children and older adults. Average BW values tend to be larger with the new method, likely reflecting ongoing trends in BW of the U.S. population, and simulated body mass indices are roughly in line with NHANES data.

2. Marginal Distributions of BW and HT

2.1. NHANES Data

For each of the NHANES cycles (2-year periods), we downloaded the age, HT, BW, and survey weights for each sampled person by merging the demographic file with the body-measurements file. We selected the variables discussed below.

Age

For 2003–2004 and 2005–2006, RIDAGEEX is the age in months at the time of examination for individuals of ages 0–84 years, and RIDAGEYR is the age in years at the time of screening for all individuals. We used RIDAGEEX to calculate the age in years for individuals under 84 (integer part of $\mathrm{RIDAGEEX}/12$) and RIDAGEYR for individuals 85 and over. We assigned the age group code “1000” to all individuals 80 and over.

For 2007–2008 and 2009–2010, RIDAGEEX is the age in months at the time of examination for individuals of ages 0–79 years, and RIDAGEYR is the age in years at the time of screening for all individuals. We used RIDAGEEX to calculate the age in years for individuals under 80 (integer part of $\mathrm{RIDAGEEX}/12$) and RIDAGEYR for individuals 80 and over. We assigned the age group code “1000” to all individuals 80 and over.

For 2009–2010 and 2011–2012, RIDEXAGM is the age in months at the time of examination for individuals of ages 0–19 years, and RIDAGEYR is the age in years at the time of screening for all individuals. We used RIDEXAGM to calculate the age in years for individuals under 20 (integer part of $\mathrm{RIDEXAGM}/12$) and RIDAGEYR for individuals 20 and over. We assigned the age group code “1000” to all individuals 80 and over.

Gender

NHANES codes gender using Males = 1 and Females = 2.

HT

For individuals of ages 2 years and older, we used the NHANES variable BMXHT, which is the standing HT (cm). For children of ages 0 or 1 years, we used the NHANES variable BMXRECUM, which is the recumbent HT (cm); for programming convenience we renamed this variable as BMXHT.

BW

For all individuals, we used the NHANES variable BMXWT, which is the BW (kg).

Survey Weight

The NHANES survey weight variable for each 2-year period is WTMEC2YR, which estimates the number of people in the U.S. population at the mid-year of the survey period represented by the sampled individual. Since the NHANES survey was designed to over-sample certain demographic groups (e.g., Mexican-Americans from 2003–2006 and Hispanics from 2006–2014), the survey weights are needed to adjust the data to represent the U.S. population.

With two exceptions, all of the analyses in this memorandum used the survey weights to adjust the data. One of these exceptions is for the histogram plots in the next sub-section, which used the survey weights rounded to the nearest integer because SAS does not allow fractional weights for those plots. A second exception is for the natural cubic spline smoothing of the parameter estimates described in Section 4; the survey weights were used in the calculations of the unsmoothed parameters but it would not have been appropriate to use them for the final smoothing step.

2.2. Histograms

Figure 2-1 and Figure 2-2 below are histograms of the BW (kg) and HT (cm; standing HT for ages 2 and over, recumbent HT for ages 0 and 1), respectively, for each gender and selected single years of age (the selected ages shown are 1, 5, 10, 15, 20, 25, 30, 40, 60, 70, and 79 years). Superimposed on each histogram are fitted normal and log-normal distributions. The calculations use the survey weights rounded to the nearest integer (making a negligible error, since the survey weights are usually several thousand). **For BW (Figure 2-1), the distributions are generally right-skewed and the log-normal distribution appears to fit the data better than the normal distribution. For HT (Figure 2-2), the distributions are almost symmetric and it is hard to distinguish the two fitted distributions on the plots**. We provide larger versions of the histograms in Figure 2-1 and Figure 2-2 in Attachment A and Attachment B, respectively.

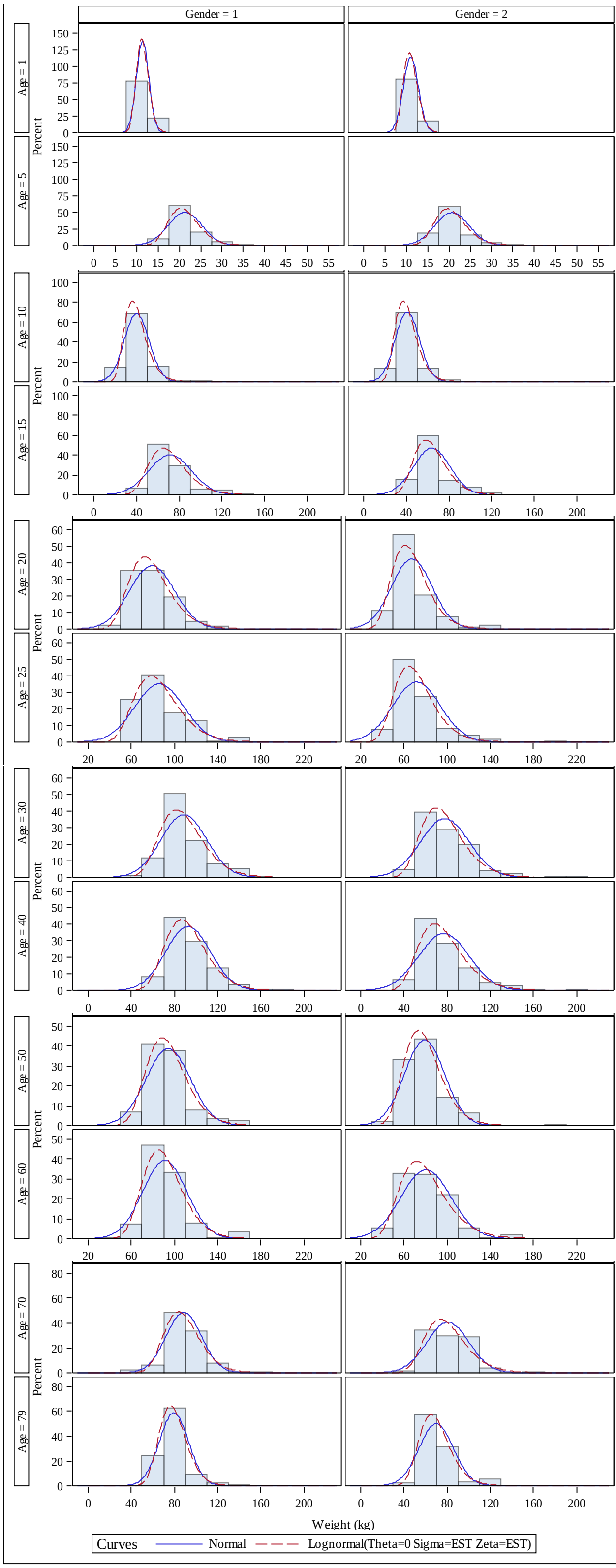


Figure 2-1. Distributions of BW

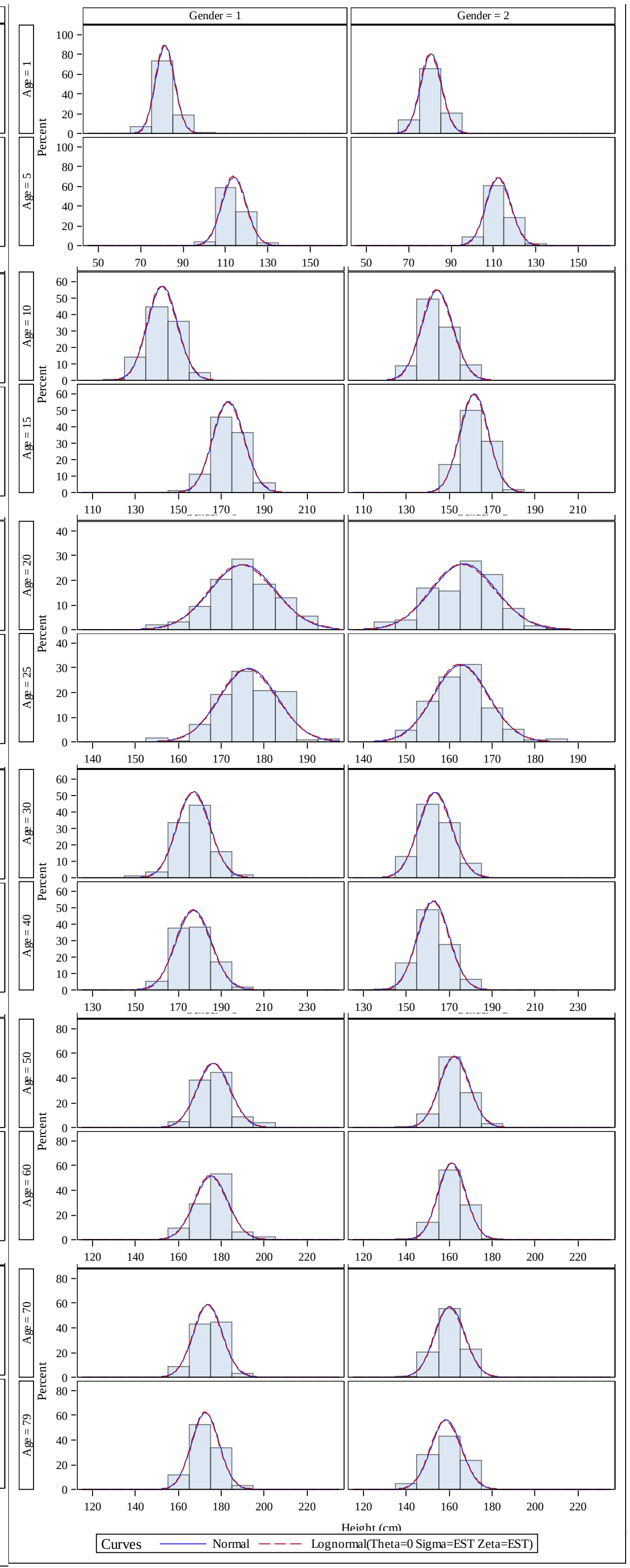


Figure 2-2. Distributions of HT

2.3. Summary Statistics

Table 1 below contains the estimated means ("Mean") and standard deviations ("Std Dev") for the BW and its natural logarithm ("Log") for each age group and gender. The row for age "1000" corresponds to ages 80 and older; the summary statistics for this group are shown for comparison purposes but are not used for the final set of distributions which are only based on the data for ages 0–79 years. Distributions are fitted separately to each combination of gender and either a single year of age from 0 to 79 years or the age group 80 years and older. We weighted the means and standard deviations across the sampled individuals using the exact survey weights.

To compare the fit of the normal and log-normal distributions, we tabulated the likelihood values. If f(x) is the probability density function for x (either a log-normal or normal distribution), then $-2\mathrm{LL} = -2 \times \Sigma\, \mathrm{SW_i} \times \log\{f(x_i)\}$, where SW_i and x_i are the survey weight and BW, respectively, for the i'th individual of the given age group and gender. (We omitted the constant term $\frac{1}{\sqrt{2\pi}}$ from f(x)). The value −2LL estimates the corresponding value of minus twice the log-likelihood for the population. Based on the likelihood method, the better of the two models (normal or log-normal) will have a lower value of −2LL; this determination is shown in the column "Best."

For the vast majority of cases, the log-normal model is preferred for BW. This pattern is consistent with the histograms shown above. Since the results of the APEX simulations should not be too sensitive to the exact ages of the modeled population, it is better to use the same distribution for all ages and genders, which suggests that **BW should be modeled as a log-normal distribution for all demographic groups**.

Table 1. Summary Statistics for BW

Age	Gender	Mean BW	Mean Log BW	Std Dev BW	Std Dev Log BW	Best	−2LL Normal	−2LL Log-Normal
0	1	7.815	2.024	1.933	0.261	Normal	1.30E+07	1.32E+07
1	1	11.443	2.429	1.451	0.126	Lognormal	1.02E+07	9.99E+06
2	1	14.130	2.640	1.850	0.126	Lognormal	1.32E+07	1.26E+07
3	1	16.162	2.773	2.436	0.139	Lognormal	1.94E+07	1.81E+07
4	1	18.693	2.915	3.152	0.157	Lognormal	2.24E+07	2.13E+07
5	1	21.347	3.045	4.002	0.172	Lognormal	2.14E+07	2.02E+07
6	1	23.789	3.149	5.344	0.191	Lognormal	2.81E+07	2.57E+07
7	1	27.870	3.298	7.526	0.234	Lognormal	3.34E+07	3.11E+07
8	1	31.112	3.407	8.244	0.241	Lognormal	3.62E+07	3.45E+07
9	1	34.679	3.513	9.531	0.249	Lognormal	3.38E+07	3.22E+07
10	1	40.133	3.656	11.645	0.263	Lognormal	3.49E+07	3.33E+07
11	1	48.057	3.832	14.351	0.280	Lognormal	3.48E+07	3.36E+07
12	1	50.746	3.894	13.498	0.252	Lognormal	3.99E+07	3.88E+07
13	1	60.002	4.060	16.631	0.256	Lognormal	4.08E+07	3.94E+07
14	1	65.258	4.143	18.467	0.259	Lognormal	5.00E+07	4.82E+07
15	1	71.356	4.234	19.846	0.255	Lognormal	4.14E+07	4.00E+07
16	1	74.894	4.289	18.367	0.226	Lognormal	4.57E+07	4.43E+07
17	1	77.237	4.317	20.101	0.235	Lognormal	4.23E+07	4.07E+07
18	1	81.164	4.363	23.222	0.248	Lognormal	4.39E+07	4.18E+07
19	1	79.636	4.350	19.629	0.229	Lognormal	4.71E+07	4.57E+07

Age	Gender	Mean BW	Mean Log BW	Std Dev BW	Std Dev Log BW	Best	−2LL Normal	−2LL Log-Normal
20	1	79.206	4.341	20.898	0.246	Lognormal	5.21E+07	5.07E+07
21	1	79.075	4.342	20.585	0.231	Lognormal	4.61E+07	4.41E+07
22	1	81.032	4.368	20.166	0.224	Lognormal	4.43E+07	4.26E+07
23	1	86.142	4.418	25.256	0.269	Lognormal	4.57E+07	4.41E+07
24	1	82.705	4.396	16.561	0.192	Lognormal	4.27E+07	4.19E+07
25	1	85.955	4.422	22.691	0.248	Lognormal	4.40E+07	4.29E+07
26	1	86.496	4.437	19.619	0.213	Lognormal	3.59E+07	3.50E+07
27	1	86.016	4.433	18.552	0.207	Lognormal	3.80E+07	3.73E+07
28	1	88.812	4.459	21.574	0.230	Lognormal	4.74E+07	4.62E+07
29	1	89.171	4.467	20.015	0.215	Lognormal	4.63E+07	4.54E+07
30	1	88.645	4.458	21.090	0.233	Lognormal	4.87E+07	4.80E+07
31	1	88.916	4.465	19.163	0.211	Lognormal	3.86E+07	3.81E+07
32	1	91.226	4.486	22.585	0.230	Lognormal	4.54E+07	4.41E+07
33	1	92.027	4.500	19.719	0.208	Lognormal	3.85E+07	3.79E+07
34	1	87.439	4.451	17.985	0.194	Lognormal	3.33E+07	3.26E+07
35	1	88.897	4.461	21.560	0.228	Lognormal	3.94E+07	3.84E+07
36	1	92.644	4.498	25.114	0.240	Lognormal	4.54E+07	4.36E+07
37	1	93.184	4.512	21.813	0.204	Lognormal	4.11E+07	3.92E+07
38	1	93.366	4.514	20.963	0.210	Lognormal	3.89E+07	3.79E+07
39	1	90.726	4.483	20.780	0.219	Lognormal	4.24E+07	4.16E+07
40	1	92.532	4.504	20.717	0.212	Lognormal	4.58E+07	4.48E+07
41	1	94.364	4.522	22.769	0.218	Lognormal	4.73E+07	4.56E+07
42	1	90.804	4.491	17.670	0.189	Lognormal	3.59E+07	3.54E+07
43	1	92.679	4.510	19.518	0.192	Lognormal	4.57E+07	4.43E+07
44	1	93.069	4.512	20.205	0.202	Lognormal	4.53E+07	4.41E+07
45	1	88.197	4.463	16.018	0.182	Lognormal	3.79E+07	3.77E+07
46	1	90.498	4.485	18.381	0.200	Lognormal	4.43E+07	4.38E+07
47	1	90.870	4.493	17.327	0.180	Lognormal	4.41E+07	4.31E+07
48	1	90.708	4.482	21.347	0.221	Lognormal	4.08E+07	3.98E+07
49	1	90.907	4.488	19.250	0.208	Lognormal	4.00E+07	3.95E+07
50	1	94.131	4.524	20.593	0.199	Lognormal	4.70E+07	4.55E+07
51	1	86.258	4.432	20.135	0.221	Lognormal	3.66E+07	3.57E+07
52	1	92.086	4.501	19.609	0.205	Lognormal	4.26E+07	4.19E+07
53	1	90.250	4.479	19.589	0.215	Lognormal	4.25E+07	4.21E+07
54	1	93.833	4.521	19.125	0.204	Lognormal	4.32E+07	4.30E+07
55	1	90.353	4.483	18.593	0.203	Lognormal	4.56E+07	4.52E+07
56	1	90.006	4.481	17.833	0.192	Lognormal	4.20E+07	4.13E+07
57	1	89.277	4.474	17.028	0.190	Lognormal	3.66E+07	3.63E+07
58	1	89.392	4.474	18.265	0.195	Lognormal	3.85E+07	3.78E+07
59	1	91.403	4.491	20.709	0.217	Lognormal	4.75E+07	4.66E+07
60	1	90.917	4.488	20.306	0.206	Lognormal	3.96E+07	3.85E+07
61	1	93.150	4.506	22.700	0.233	Lognormal	3.16E+07	3.10E+07
62	1	90.499	4.487	18.053	0.192	Lognormal	3.11E+07	3.06E+07
63	1	91.326	4.486	23.270	0.234	Lognormal	3.80E+07	3.68E+07
64	1	89.615	4.467	23.395	0.230	Lognormal	3.13E+07	3.00E+07
65	1	91.754	4.493	20.739	0.229	Lognormal	3.88E+07	3.86E+07
66	1	89.407	4.471	18.910	0.210	Lognormal	2.57E+07	2.55E+07
67	1	90.274	4.482	18.677	0.207	Lognormal	1.96E+07	1.95E+07
68	1	88.174	4.447	22.562	0.256	Lognormal	2.67E+07	2.64E+07
69	1	88.345	4.461	17.487	0.204	Normal	2.12E+07	2.13E+07

Age	Gender	Mean BW	Mean Log BW	Std Dev BW	Std Dev Log BW	Best	−2LL Normal	−2LL Log-Normal
70	1	88.508	4.465	16.451	0.190	Normal	2.32E+07	2.32E+07
71	1	86.951	4.442	19.122	0.218	Lognormal	1.23E+07	1.22E+07
72	1	85.011	4.427	14.707	0.184	Normal	2.07E+07	2.10E+07
73	1	82.985	4.401	16.298	0.189	Lognormal	1.48E+07	1.45E+07
74	1	87.057	4.452	15.113	0.172	Lognormal	1.71E+07	1.69E+07
75	1	84.965	4.418	18.599	0.219	Lognormal	1.54E+07	1.53E+07
76	1	84.242	4.418	15.364	0.173	Lognormal	1.45E+07	1.42E+07
77	1	87.413	4.457	14.289	0.166	Normal	1.19E+07	1.19E+07
78	1	86.227	4.437	17.646	0.199	Lognormal	1.08E+07	1.06E+07
79	1	79.399	4.361	13.595	0.160	Lognormal	7.74E+06	7.54E+06
1000	1	79.526	4.360	14.305	0.182	Lognormal	7.06E+07	7.04E+07
0	2	7.370	1.963	1.848	0.270	Normal	1.19E+07	1.23E+07
1	2	11.090	2.394	1.754	0.152	Lognormal	1.14E+07	1.09E+07
2	2	13.219	2.573	1.838	0.133	Lognormal	1.43E+07	1.36E+07
3	2	15.640	2.739	2.510	0.145	Lognormal	1.70E+07	1.56E+07
4	2	18.059	2.879	3.247	0.168	Lognormal	2.17E+07	2.06E+07
5	2	20.679	3.012	4.027	0.181	Lognormal	2.12E+07	2.02E+07
6	2	23.793	3.147	5.253	0.205	Lognormal	2.36E+07	2.26E+07
7	2	26.881	3.261	7.211	0.238	Lognormal	2.92E+07	2.75E+07
8	2	32.029	3.433	9.019	0.253	Lognormal	2.99E+07	2.84E+07
9	2	36.699	3.566	10.701	0.264	Lognormal	3.46E+07	3.30E+07
10	2	41.050	3.681	11.396	0.256	Lognormal	3.30E+07	3.17E+07
11	2	47.362	3.818	13.982	0.278	Lognormal	4.43E+07	4.29E+07
12	2	54.672	3.963	15.597	0.273	Lognormal	4.31E+07	4.20E+07
13	2	56.288	4.000	14.933	0.242	Lognormal	3.57E+07	3.44E+07
14	2	59.807	4.069	13.215	0.209	Lognormal	4.03E+07	3.92E+07
15	2	63.838	4.126	16.980	0.240	Lognormal	4.48E+07	4.30E+07
16	2	64.978	4.140	18.345	0.251	Lognormal	4.31E+07	4.12E+07
17	2	65.573	4.151	18.055	0.244	Lognormal	4.11E+07	3.92E+07
18	2	67.681	4.177	20.459	0.263	Lognormal	4.15E+07	3.94E+07
19	2	68.713	4.193	20.005	0.266	Lognormal	3.53E+07	3.40E+07
20	2	67.242	4.175	18.889	0.250	Lognormal	4.92E+07	4.70E+07
21	2	68.518	4.194	18.688	0.253	Lognormal	4.11E+07	3.98E+07
22	2	73.589	4.263	21.062	0.257	Lognormal	4.77E+07	4.57E+07
23	2	73.890	4.269	19.737	0.258	Lognormal	4.70E+07	4.61E+07
24	2	74.087	4.270	20.804	0.259	Lognormal	3.92E+07	3.79E+07
25	2	71.664	4.235	22.042	0.261	Lognormal	4.91E+07	4.63E+07
26	2	74.947	4.278	22.693	0.268	Lognormal	4.46E+07	4.26E+07
27	2	76.495	4.300	21.714	0.272	Lognormal	4.47E+07	4.37E+07
28	2	76.115	4.293	22.452	0.274	Lognormal	4.54E+07	4.40E+07
29	2	76.079	4.305	17.674	0.234	Lognormal	3.79E+07	3.77E+07
30	2	77.839	4.318	22.534	0.262	Lognormal	4.31E+07	4.14E+07
31	2	77.715	4.316	22.610	0.264	Lognormal	5.03E+07	4.84E+07
32	2	79.498	4.331	26.226	0.289	Lognormal	3.77E+07	3.59E+07
33	2	80.160	4.353	21.345	0.243	Lognormal	4.10E+07	3.96E+07
34	2	79.954	4.341	24.352	0.278	Lognormal	4.26E+07	4.11E+07
35	2	76.240	4.309	17.070	0.221	Lognormal	3.04E+07	3.01E+07
36	2	76.700	4.304	22.247	0.259	Lognormal	5.09E+07	4.88E+07
37	2	79.289	4.333	23.794	0.276	Lognormal	4.06E+07	3.92E+07
38	2	79.992	4.354	19.236	0.236	Lognormal	4.41E+07	4.36E+07

Age	Gender	Mean BW	Mean Log BW	Std Dev BW	Std Dev Log BW	Best	−2LL Normal	−2LL Log-Normal
39	2	76.566	4.305	21.337	0.251	Lognormal	4.81E+07	4.62E+07
40	2	76.974	4.303	23.274	0.279	Lognormal	4.61E+07	4.46E+07
41	2	76.441	4.301	21.868	0.260	Lognormal	4.68E+07	4.51E+07
42	2	76.145	4.298	20.347	0.264	Lognormal	4.63E+07	4.57E+07
43	2	76.903	4.311	20.853	0.243	Lognormal	4.84E+07	4.65E+07
44	2	75.614	4.290	22.250	0.260	Lognormal	4.77E+07	4.55E+07
45	2	75.209	4.290	20.478	0.238	Lognormal	4.98E+07	4.74E+07
46	2	79.677	4.348	21.220	0.240	Lognormal	3.92E+07	3.77E+07
47	2	80.825	4.360	21.865	0.249	Lognormal	4.76E+07	4.60E+07
48	2	78.180	4.324	21.616	0.260	Lognormal	4.68E+07	4.56E+07
49	2	78.804	4.338	19.602	0.240	Lognormal	4.61E+07	4.53E+07
50	2	79.090	4.345	18.574	0.221	Lognormal	5.30E+07	5.17E+07
51	2	77.540	4.320	20.179	0.244	Lognormal	4.67E+07	4.54E+07
52	2	73.712	4.267	20.579	0.252	Lognormal	5.12E+07	4.93E+07
53	2	77.885	4.325	19.474	0.243	Lognormal	3.77E+07	3.70E+07
54	2	81.799	4.368	23.266	0.262	Lognormal	4.49E+07	4.35E+07
55	2	81.660	4.364	23.736	0.270	Lognormal	4.30E+07	4.17E+07
56	2	78.463	4.332	19.938	0.245	Lognormal	5.21E+07	5.11E+07
57	2	77.206	4.320	19.414	0.225	Lognormal	4.11E+07	3.95E+07
58	2	82.906	4.372	25.218	0.306	Lognormal	3.27E+07	3.24E+07
59	2	75.924	4.305	17.461	0.223	Lognormal	4.32E+07	4.25E+07
60	2	80.438	4.349	23.023	0.276	Lognormal	4.03E+07	3.95E+07
61	2	81.177	4.374	17.290	0.215	Lognormal	4.17E+07	4.15E+07
62	2	81.189	4.373	18.224	0.216	Lognormal	3.11E+07	3.05E+07
63	2	74.279	4.282	17.151	0.229	Lognormal	3.96E+07	3.92E+07
64	2	78.502	4.333	20.131	0.243	Lognormal	4.00E+07	3.91E+07
65	2	74.259	4.284	16.038	0.219	Lognormal	3.21E+07	3.20E+07
66	2	76.788	4.320	15.800	0.207	Lognormal	2.52E+07	2.51E+07
67	2	77.607	4.318	20.286	0.259	Lognormal	2.64E+07	2.61E+07
68	2	71.134	4.237	17.438	0.232	Lognormal	2.51E+07	2.45E+07
69	2	74.826	4.288	16.942	0.237	Normal	2.21E+07	2.22E+07
70	2	80.651	4.361	19.520	0.243	Lognormal	2.93E+07	2.91E+07
71	2	77.613	4.318	20.636	0.259	Lognormal	2.55E+07	2.51E+07
72	2	75.780	4.295	19.888	0.254	Lognormal	2.38E+07	2.33E+07
73	2	76.332	4.307	18.416	0.234	Lognormal	2.22E+07	2.18E+07
74	2	73.923	4.280	16.136	0.216	Lognormal	2.19E+07	2.16E+07
75	2	73.693	4.276	15.862	0.222	Normal	1.45E+07	1.45E+07
76	2	77.133	4.324	16.505	0.209	Lognormal	1.55E+07	1.53E+07
77	2	73.587	4.270	18.167	0.238	Lognormal	1.30E+07	1.27E+07
78	2	72.360	4.258	16.423	0.216	Lognormal	1.29E+07	1.26E+07
79	2	69.868	4.224	15.927	0.208	Lognormal	1.45E+07	1.40E+07
1000	2	64.634	4.148	13.273	0.205	Lognormal	1.13E+08	1.12E+08

Note: Age 1000 = 80 years or older.

Table 2 below is the same as Table 1 above but for HT. In this case, the preferred distribution is less consistent since 64 percent of the HT cases have “Normal” for the “Best” distribution and 36 percent of the cases have “Lognormal.” The histograms also did not show a strong preference for one of those two distributions. Since the results of the APEX simulations should not be too sensitive to the exact ages of the modeled population, it is better to use the same distribution for all ages and genders, which suggests that **HT should be modeled as a normal distribution for all demographic groups**.

Table 2. Summary Statistics for HT

Age	Gender	Mean HT	Mean Log HT	Std Dev HT	Std Dev Log HT	Best	−2LL Normal	−2LL Log-Normal
0	1	66.348	4.190	6.538	0.101	Normal	2.66E+07	2.68E+07
1	1	81.551	4.400	4.495	0.055	Lognormal	2.33E+07	2.32E+07
2	1	91.720	4.518	4.508	0.049	Normal	2.32E+07	2.32E+07
3	1	98.932	4.593	4.763	0.048	Normal	2.86E+07	2.86E+07
4	1	106.749	4.669	4.795	0.045	Lognormal	2.81E+07	2.81E+07
5	1	114.047	4.735	5.750	0.050	Lognormal	2.55E+07	2.54E+07
6	1	119.584	4.783	5.647	0.047	Normal	2.87E+07	2.88E+07
7	1	126.274	4.837	6.172	0.049	Normal	3.08E+07	3.08E+07
8	1	131.387	4.877	6.487	0.050	Normal	3.28E+07	3.28E+07
9	1	137.145	4.920	6.989	0.051	Lognormal	3.00E+07	2.99E+07
10	1	142.600	4.959	6.965	0.049	Normal	2.88E+07	2.89E+07
11	1	150.274	5.011	8.441	0.056	Lognormal	2.89E+07	2.88E+07
12	1	155.594	5.046	7.455	0.048	Lognormal	3.23E+07	3.23E+07
13	1	163.822	5.097	8.320	0.051	Normal	3.23E+07	3.24E+07
14	1	168.833	5.128	7.825	0.047	Normal	3.74E+07	3.75E+07
15	1	173.395	5.155	7.224	0.042	Normal	2.94E+07	2.95E+07
16	1	174.662	5.162	6.608	0.038	Normal	3.20E+07	3.21E+07
17	1	175.483	5.166	8.067	0.046	Normal	3.13E+07	3.13E+07
18	1	175.871	5.169	7.309	0.042	Normal	3.00E+07	3.00E+07
19	1	176.655	5.173	7.524	0.043	Lognormal	3.41E+07	3.41E+07
20	1	175.034	5.164	7.566	0.044	Normal	3.72E+07	3.73E+07
21	1	176.763	5.174	8.403	0.048	Normal	3.49E+07	3.50E+07
22	1	176.195	5.171	6.516	0.037	Lognormal	3.00E+07	3.00E+07
23	1	174.777	5.162	8.261	0.047	Lognormal	3.20E+07	3.19E+07
24	1	176.734	5.174	7.498	0.042	Lognormal	3.25E+07	3.24E+07
25	1	176.400	5.172	6.713	0.038	Normal	2.92E+07	2.93E+07
26	1	176.482	5.172	6.841	0.039	Normal	2.50E+07	2.51E+07
27	1	176.625	5.173	6.835	0.039	Normal	2.70E+07	2.70E+07
28	1	177.668	5.179	7.591	0.043	Normal	3.35E+07	3.35E+07
29	1	176.629	5.173	7.984	0.045	Lognormal	3.41E+07	3.40E+07
30	1	177.154	5.176	7.644	0.044	Normal	3.48E+07	3.49E+07
31	1	176.424	5.172	6.393	0.036	Normal	2.63E+07	2.63E+07
32	1	176.506	5.172	8.069	0.046	Normal	3.25E+07	3.26E+07
33	1	177.685	5.179	7.686	0.043	Lognormal	2.81E+07	2.81E+07
34	1	176.909	5.175	7.629	0.043	Normal	2.49E+07	2.49E+07
35	1	175.465	5.166	8.162	0.047	Normal	2.87E+07	2.88E+07
36	1	175.886	5.169	7.555	0.043	Normal	3.08E+07	3.08E+07
37	1	176.134	5.170	7.465	0.043	Normal	2.88E+07	2.88E+07
38	1	176.737	5.174	7.627	0.043	Normal	2.78E+07	2.78E+07
39	1	176.688	5.173	8.195	0.047	Normal	3.13E+07	3.14E+07
40	1	177.188	5.176	8.246	0.046	Lognormal	3.41E+07	3.40E+07
41	1	177.129	5.176	8.370	0.047	Normal	3.42E+07	3.43E+07
42	1	175.377	5.166	7.477	0.043	Lognormal	2.67E+07	2.67E+07
43	1	177.690	5.179	7.330	0.041	Lognormal	3.28E+07	3.28E+07

Age	Gender	Mean HT	Mean Log HT	Std Dev HT	Std Dev Log HT	Best	−2LL Normal	−2LL Log-Normal
44	1	176.112	5.170	7.903	0.045	Lognormal	3.32E+07	3.31E+07
45	1	174.981	5.164	7.396	0.042	Normal	2.89E+07	2.90E+07
46	1	176.634	5.173	6.562	0.038	Normal	3.09E+07	3.10E+07
47	1	175.600	5.167	6.753	0.038	Lognormal	3.17E+07	3.17E+07
48	1	176.122	5.170	7.434	0.043	Normal	2.87E+07	2.88E+07
49	1	177.033	5.176	6.807	0.039	Normal	2.78E+07	2.79E+07
50	1	176.496	5.172	7.690	0.043	Lognormal	3.39E+07	3.38E+07
51	1	174.912	5.163	7.901	0.045	Lognormal	2.69E+07	2.69E+07
52	1	176.530	5.173	6.804	0.039	Normal	2.96E+07	2.96E+07
53	1	176.744	5.174	7.201	0.041	Lognormal	3.02E+07	3.02E+07
54	1	176.288	5.171	7.453	0.042	Normal	3.16E+07	3.16E+07
55	1	175.405	5.166	6.225	0.035	Lognormal	3.10E+07	3.10E+07
56	1	176.729	5.174	7.468	0.043	Normal	3.09E+07	3.10E+07
57	1	175.733	5.168	8.368	0.048	Normal	2.88E+07	2.89E+07
58	1	176.871	5.174	8.038	0.046	Normal	2.93E+07	2.93E+07
59	1	176.603	5.173	6.358	0.036	Normal	3.16E+07	3.17E+07
60	1	175.322	5.166	7.743	0.044	Lognormal	2.90E+07	2.89E+07
61	1	175.231	5.165	7.553	0.044	Normal	2.20E+07	2.20E+07
62	1	174.979	5.164	7.231	0.042	Normal	2.27E+07	2.28E+07
63	1	177.680	5.179	8.229	0.046	Lognormal	2.69E+07	2.69E+07
64	1	173.887	5.158	7.268	0.042	Normal	2.13E+07	2.14E+07
65	1	175.770	5.168	7.209	0.042	Normal	2.72E+07	2.73E+07
66	1	175.376	5.166	8.807	0.051	Normal	2.00E+07	2.01E+07
67	1	173.978	5.158	6.767	0.039	Lognormal	1.38E+07	1.38E+07
68	1	174.040	5.159	6.660	0.039	Normal	1.81E+07	1.82E+07
69	1	173.767	5.157	8.313	0.048	Normal	1.66E+07	1.66E+07
70	1	173.764	5.157	6.780	0.039	Normal	1.69E+07	1.69E+07
71	1	171.952	5.146	7.098	0.041	Lognormal	8.79E+06	8.75E+06
72	1	173.617	5.156	7.523	0.044	Normal	1.64E+07	1.64E+07
73	1	171.815	5.145	7.548	0.044	Normal	1.14E+07	1.14E+07
74	1	173.762	5.157	6.224	0.036	Lognormal	1.23E+07	1.22E+07
75	1	172.609	5.150	7.212	0.042	Lognormal	1.12E+07	1.12E+07
76	1	172.734	5.151	6.328	0.037	Lognormal	1.05E+07	1.05E+07
77	1	172.442	5.149	7.440	0.043	Normal	9.47E+06	9.48E+06
78	1	174.156	5.159	7.499	0.043	Normal	7.98E+06	7.98E+06
79	1	172.635	5.150	6.417	0.037	Lognormal	5.87E+06	5.86E+06
1000	1	171.292	5.143	6.915	0.041	Normal	5.32E+07	5.32E+07
0	2	64.997	4.169	6.275	0.100	Normal	2.50E+07	2.52E+07
1	2	80.615	4.388	4.947	0.062	Normal	2.25E+07	2.25E+07
2	2	89.528	4.493	4.204	0.046	Lognormal	2.50E+07	2.49E+07
3	2	98.281	4.587	4.248	0.044	Normal	2.29E+07	2.30E+07
4	2	105.404	4.657	4.857	0.046	Normal	2.69E+07	2.70E+07
5	2	112.415	4.721	5.787	0.052	Lognormal	2.53E+07	2.53E+07
6	2	118.957	4.778	5.654	0.048	Normal	2.44E+07	2.44E+07
7	2	124.658	4.824	5.843	0.047	Lognormal	2.68E+07	2.67E+07
8	2	131.786	4.880	6.950	0.052	Lognormal	2.70E+07	2.69E+07
9	2	137.722	4.924	6.500	0.047	Lognormal	2.86E+07	2.86E+07
10	2	144.426	4.971	7.298	0.050	Lognormal	2.80E+07	2.79E+07
11	2	150.574	5.013	7.670	0.052	Normal	3.58E+07	3.60E+07
12	2	156.583	5.052	7.295	0.047	Normal	3.30E+07	3.31E+07

Age	Gender	Mean HT	Mean Log HT	Std Dev HT	Std Dev Log HT	Best	−2LL Normal	−2LL Log-Normal
13	2	158.923	5.068	6.149	0.039	Lognormal	2.58E+07	2.58E+07
14	2	160.849	5.080	6.429	0.040	Normal	3.09E+07	3.09E+07
15	2	161.704	5.085	6.674	0.042	Normal	3.22E+07	3.23E+07
16	2	162.002	5.087	6.219	0.038	Lognormal	2.94E+07	2.94E+07
17	2	162.805	5.092	6.661	0.041	Normal	2.95E+07	2.95E+07
18	2	162.208	5.088	6.344	0.039	Lognormal	2.77E+07	2.77E+07
19	2	163.320	5.095	6.174	0.038	Normal	2.35E+07	2.35E+07
20	2	163.411	5.095	7.485	0.046	Normal	3.59E+07	3.60E+07
21	2	161.858	5.086	6.643	0.041	Lognormal	2.87E+07	2.86E+07
22	2	162.038	5.087	6.058	0.037	Lognormal	3.09E+07	3.09E+07
23	2	161.916	5.086	7.447	0.046	Normal	3.38E+07	3.39E+07
24	2	162.774	5.091	7.195	0.044	Lognormal	2.74E+07	2.73E+07
25	2	162.763	5.092	6.405	0.039	Lognormal	3.22E+07	3.21E+07
26	2	163.198	5.094	6.312	0.039	Normal	2.90E+07	2.91E+07
27	2	163.593	5.096	7.471	0.046	Normal	3.14E+07	3.14E+07
28	2	163.380	5.095	6.569	0.040	Normal	2.99E+07	3.00E+07
29	2	162.909	5.093	5.527	0.034	Normal	2.49E+07	2.49E+07
30	2	163.515	5.096	7.695	0.047	Normal	3.03E+07	3.03E+07
31	2	164.013	5.099	6.712	0.041	Normal	3.34E+07	3.34E+07
32	2	163.674	5.097	7.194	0.044	Normal	2.48E+07	2.48E+07
33	2	163.856	5.098	6.710	0.041	Normal	2.77E+07	2.78E+07
34	2	163.344	5.095	7.496	0.046	Lognormal	2.90E+07	2.90E+07
35	2	163.531	5.096	6.544	0.041	Normal	2.17E+07	2.18E+07
36	2	163.211	5.094	7.656	0.047	Normal	3.58E+07	3.58E+07
37	2	164.099	5.100	6.902	0.043	Normal	2.69E+07	2.70E+07
38	2	162.956	5.092	7.860	0.048	Lognormal	3.27E+07	3.26E+07
39	2	162.702	5.091	7.675	0.047	Normal	3.44E+07	3.44E+07
40	2	162.678	5.091	7.397	0.045	Lognormal	3.16E+07	3.16E+07
41	2	161.638	5.085	6.643	0.041	Lognormal	3.13E+07	3.12E+07
42	2	163.154	5.094	7.131	0.043	Lognormal	3.28E+07	3.27E+07
43	2	162.756	5.091	6.773	0.042	Normal	3.30E+07	3.30E+07
44	2	162.821	5.092	6.921	0.043	Normal	3.22E+07	3.23E+07
45	2	162.737	5.092	5.720	0.035	Normal	3.19E+07	3.19E+07
46	2	162.146	5.087	7.539	0.047	Normal	2.79E+07	2.80E+07
47	2	163.495	5.096	7.326	0.045	Normal	3.31E+07	3.31E+07
48	2	163.566	5.096	6.311	0.039	Normal	3.07E+07	3.08E+07
49	2	162.858	5.092	6.338	0.039	Normal	3.11E+07	3.13E+07
50	2	162.498	5.090	6.919	0.043	Normal	3.76E+07	3.77E+07
51	2	162.610	5.091	5.990	0.037	Normal	3.06E+07	3.07E+07
52	2	161.654	5.084	7.879	0.051	Normal	3.73E+07	3.80E+07
53	2	163.379	5.095	6.657	0.041	Normal	2.60E+07	2.61E+07
54	2	162.049	5.087	7.027	0.043	Lognormal	3.02E+07	3.01E+07
55	2	162.694	5.091	6.633	0.041	Normal	2.81E+07	2.81E+07
56	2	162.638	5.091	6.787	0.041	Lognormal	3.60E+07	3.59E+07
57	2	160.512	5.077	7.084	0.044	Lognormal	2.92E+07	2.91E+07
58	2	160.963	5.080	7.017	0.044	Normal	2.15E+07	2.15E+07
59	2	160.849	5.080	6.991	0.043	Lognormal	3.15E+07	3.14E+07
60	2	161.262	5.082	6.422	0.040	Normal	2.62E+07	2.63E+07
61	2	163.010	5.093	7.148	0.044	Lognormal	3.07E+07	3.07E+07
62	2	160.395	5.077	6.512	0.041	Lognormal	2.17E+07	2.17E+07

Age	Gender	Mean HT	Mean Log HT	Std Dev HT	Std Dev Log HT	Best	−2LL Normal	−2LL Log-Normal
63	2	161.629	5.084	6.589	0.041	Lognormal	2.83E+07	2.82E+07
64	2	160.269	5.076	6.028	0.038	Normal	2.69E+07	2.70E+07
65	2	161.070	5.081	6.539	0.040	Lognormal	2.33E+07	2.32E+07
66	2	159.425	5.071	5.689	0.036	Normal	1.74E+07	1.75E+07
67	2	160.241	5.076	6.903	0.043	Lognormal	1.83E+07	1.83E+07
68	2	158.931	5.067	7.056	0.045	Normal	1.82E+07	1.83E+07
69	2	159.863	5.073	6.687	0.043	Normal	1.59E+07	1.60E+07
70	2	160.263	5.076	6.986	0.044	Normal	2.07E+07	2.07E+07
71	2	159.678	5.072	7.340	0.046	Normal	1.80E+07	1.80E+07
72	2	158.699	5.066	6.225	0.039	Lognormal	1.59E+07	1.59E+07
73	2	159.618	5.072	7.187	0.045	Normal	1.61E+07	1.61E+07
74	2	159.042	5.068	6.425	0.040	Lognormal	1.57E+07	1.57E+07
75	2	158.332	5.064	7.461	0.047	Normal	1.11E+07	1.11E+07
76	2	159.769	5.073	5.740	0.036	Normal	1.05E+07	1.05E+07
77	2	158.186	5.063	5.841	0.037	Normal	8.57E+06	8.58E+06
78	2	158.001	5.062	7.098	0.045	Normal	9.55E+06	9.57E+06
79	2	158.586	5.065	7.097	0.045	Normal	1.12E+07	1.12E+07
1000	2	155.746	5.047	6.564	0.042	Normal	8.63E+07	8.64E+07

Note: Age 1000 = 80 years or older.

For an overall comparison, we calculated the values of -2LL for the entire population ages 0–79 years by summing the values of -2LL across all ages and genders. For BW, the -2LL totals were 5.91×10^9 for the normal distribution and 5.75×10^9 for the log-normal distribution—again supporting the log-normal distribution. For HT, the -2LL totals were 4.42×10^9 for the normal distribution and 4.43×10^9 for the log-normal distribution, which provides some small support for the normal distribution. The unrounded summary statistics from Table 1 and Table 2 above are shown in the tabs “Mean”, “Weights”, and “HTs” of the accompanying Excel file “means.2009 to 2014.102016.xlsx”; the tab “Read Me” gives the content and formats for each tab.

To summarize these results, the recommended distributions are a normal distribution for HTs and a log-normal distribution for BWs. **The parameters vary by age (in years) and gender. The same conclusion was reached by Brainard and Burmaster (1992)[3]**. Note that in 2002, the CDC developed growth charts for children by fitting more complicated Box-Cox models to earlier NHANES data.[4] The Box-Cox model uses a power of the normal distribution, which tends to a log-normal distribution when the power tends to zero. Those approaches would be harder to implement for APEX, particularly when developing joint distributions for BW and HT.

3. Joint Distributions for BW and HT

The conclusion from Section 2 was that, for each age and gender, we should model BW by a log-normal distribution and HT by a normal distribution. To fit a joint distribution, it is important to

[3] Brainard, J., Burmaster, D.E. “Bivariate distributions for height and weight of men and women in the United States”. *Risk Analysis* 1992, 12(2) 267-275.

[4] http://www.cdc.gov/growthcharts/cdc_charts.htm

realize that HT and BW are not independent. Therefore, **we fit the joint distribution of HT and BW by assuming that the HT and the logarithm of the BW have a bivariate normal distribution.** Table 1 and Table 2 above contain the means and standard deviations of the HT and the logarithm of the BW. Table 3 below contains the correlations between the HT and the logarithm of the BW, calculated using the survey weights. The “Mean” tab of the accompanying Excel file “means.2009 to 2014.102016.xlsx” contains the unrounded values of the correlation coefficient.

Table 3. Correlation Between Log BW and HT

Age	Gender	Correlation Between Log BW and HT	Age	Gender	Correlation Between Log BW and HT
0	1	0.934	0	2	0.933
1	1	0.804	1	2	0.789
2	1	0.751	2	2	0.765
3	1	0.742	3	2	0.733
4	1	0.755	4	2	0.761
5	1	0.741	5	2	0.744
6	1	0.758	6	2	0.734
7	1	0.706	7	2	0.753
8	1	0.768	8	2	0.720
9	1	0.721	9	2	0.676
10	1	0.685	10	2	0.729
11	1	0.697	11	2	0.606
12	1	0.671	12	2	0.558
13	1	0.563	13	2	0.391
14	1	0.585	14	2	0.344
15	1	0.485	15	2	0.461
16	1	0.430	16	2	0.364
17	1	0.416	17	2	0.359
18	1	0.451	18	2	0.228
19	1	0.312	19	2	0.227
20	1	0.504	20	2	0.294
21	1	0.426	21	2	0.397
22	1	0.299	22	2	0.086
23	1	0.423	23	2	0.294
24	1	0.391	24	2	0.236
25	1	0.388	25	2	0.288
26	1	0.396	26	2	0.325
27	1	0.515	27	2	0.356
28	1	0.337	28	2	0.354
29	1	0.174	29	2	0.269
30	1	0.597	30	2	0.269
31	1	0.298	31	2	0.212
32	1	0.482	32	2	0.248
33	1	0.528	33	2	0.269
34	1	0.292	34	2	0.283
35	1	0.279	35	2	0.200
36	1	0.519	36	2	0.362
37	1	0.434	37	2	0.391
38	1	0.453	38	2	0.328
39	1	0.373	39	2	0.396

Age	Gender	Correlation Between Log BW and HT	Age	Gender	Correlation Between Log BW and HT
40	1	0.546	40	2	0.302
41	1	0.357	41	2	0.367
42	1	0.339	42	2	0.300
43	1	0.367	43	2	0.233
44	1	0.470	44	2	0.301
45	1	0.453	45	2	0.240
46	1	0.227	46	2	0.245
47	1	0.405	47	2	0.254
48	1	0.357	48	2	0.042
49	1	0.496	49	2	0.262
50	1	0.590	50	2	0.248
51	1	0.534	51	2	0.167
52	1	0.338	52	2	0.347
53	1	0.510	53	2	0.260
54	1	0.441	54	2	0.235
55	1	0.363	55	2	0.178
56	1	0.292	56	2	0.115
57	1	0.437	57	2	0.301
58	1	0.324	58	2	0.287
59	1	0.472	59	2	0.266
60	1	0.380	60	2	0.414
61	1	0.387	61	2	0.380
62	1	0.475	62	2	0.266
63	1	0.520	63	2	0.310
64	1	0.534	64	2	0.248
65	1	0.372	65	2	0.240
66	1	0.408	66	2	0.331
67	1	0.627	67	2	0.351
68	1	0.490	68	2	0.300
69	1	0.510	69	2	0.287
70	1	0.434	70	2	0.257
71	1	0.413	71	2	0.275
72	1	0.527	72	2	0.262
73	1	0.578	73	2	0.302
74	1	0.220	74	2	0.237
75	1	0.503	75	2	0.083
76	1	0.161	76	2	0.297
77	1	0.400	77	2	0.248
78	1	0.524	78	2	0.292
79	1	0.195	79	2	0.461
1000	1	0.491	1000	2	0.419

Note: Age 1000 = 80 years or older.

Figure 3-1 below illustrates the fitted joint distributions for selected ages (5, 15, 25, 40, 60, and 79 years) and both genders. Each data point shows the HT and the logarithm of the BW for a single NHANES subject. The red prediction ellipse includes 95 percent of the fitted joint distribution (which is not necessarily 95 percent of the sampled data). The blue prediction ellipse includes 80 percent of the fitted joint distribution (which is not necessarily 80 percent of the sampled data). The ellipses and correlations were computed using the survey weights, even though there is only a single point shown for each NHANES subject. The elliptical shapes of the scatter plot data support the use of a bivariate normal distribution with a non-zero correlation. A zero correlation would imply that HT and BW are independent. We provide larger versions of the plots in Figure 3-1 in Attachment C.

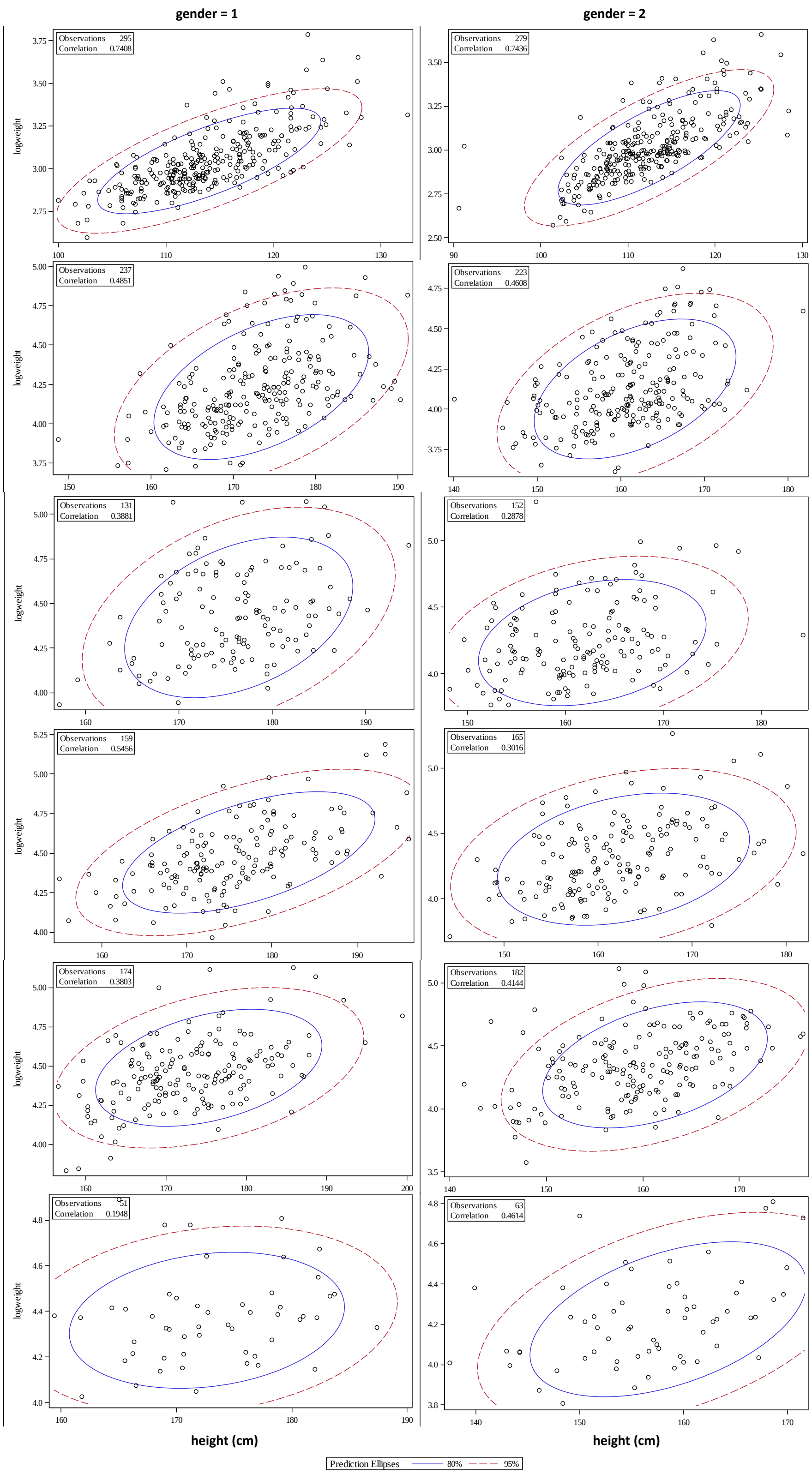


Figure 3-1. Scatter Plots of Log BW versus HT, Years 2009–2014

4. Smoothing the Parameters

4.1. Smooth Parameters Using Natural Cubic Spline

The last step for fitting the joint distributions of BW and HT is to **smooth the parameter values to make them continuous functions of the age rather than varying discontinuously**. Otherwise, a small change in the age of one of the simulated persons can lead to a large change in the simulated distribution of that person's HT and BW and thus other exposure parameters. The five parameters for each age and gender are

- mean log BW,
- standard deviation log BW,
- mean HT,
- standard deviation HT, and
- correlation.

Figure 4-1 below illustrates how the five parameters vary by age for the same gender. Also shown are the smoothed curves created with a natural cubic spline, without applying any weighting. For each parameter, we chose the same set of eight knots for the spline function: 0, 10, 20, 30, 40, 50, 60, and 70. Between each two consecutive knots, we fitted a cubic polynomial so that the curve and its first two derivatives are continuous at the knot. For values above 70, we fitted a straight line so that the curve and its first derivative are continuous at 70. (A similar linear curve applies below zero but those values are not needed since age cannot be negative**). The straight line fitted to ages 70 and above is used to extrapolate the parameter values up to age 100.** We provide larger versions of the plots in Figure 4-1 in Attachment D.

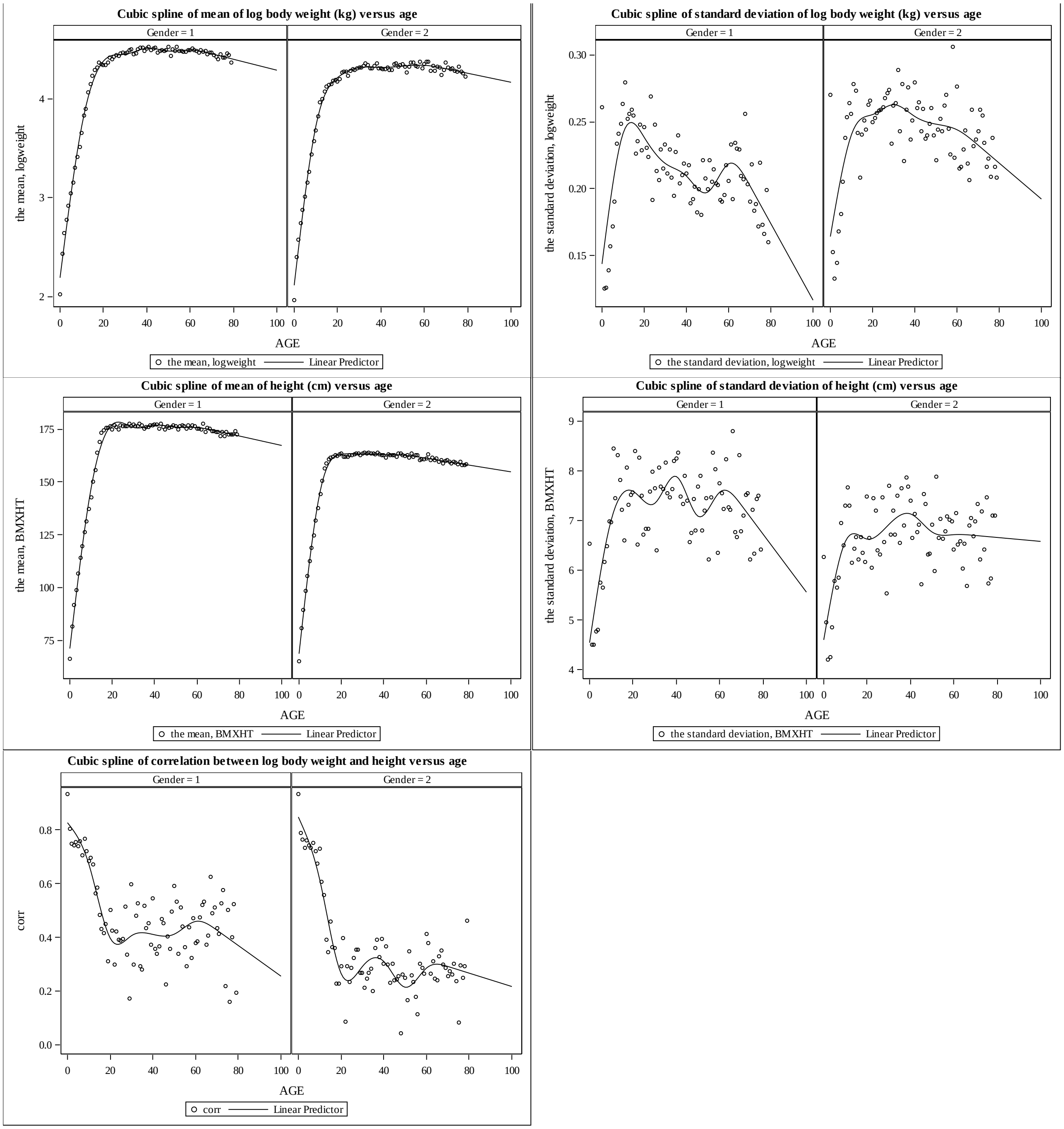


Figure 4-1. Unsmoothed and Smoothed Values for the Five Joint-distribution Parameters, Years 2009–2014

4.2. Final Parameter Values

Table 4 below, and the tab “Parameters” of the accompanying Excel file “means.2009 to 2014.102016.xlsx”, contain the unsmoothed and smoothed parameter values. The values in the Excel file should be used for APEX implementation, as they are provided with full numeric precision there.

For simulating the joint distribution of BW and HT in APEX, we propose the following approach.

First, **simulate the values of log BW from a normal distribution**. We show the mean and standard deviation of the log BW for each age and gender in the "SMOOTHED" columns of Table 4. **Truncate the distribution** at the lower and upper bounds as shown in the "BOUNDS FOR LOG BW" columns, which we calculated as

$$\text{BOUNDS FOR LOG BW} = \text{Mean Log BW} \pm (z_{0.99} \times \text{Std Dev Log BW}).$$

$z_{0.99}$ is the 99th percentile of a standard normal distribution. **Resampling should be done**, so that a new value should be selected if the simulated value is outside these bounds. Thus, the probability of being outside these two bounds is 0.02. Let w be the simulated value of log BW.

Second, **simulate the values of HT from the conditional distribution of HT given that the log of the BW is w**. The simulated value of HT is

$$\text{Simulated HT} = \text{mh} + \left(\text{sh} \times \text{corr} \times \frac{\text{w}-\text{mw}}{\text{sw}}\right) + \left(\text{sh} \times \sqrt{1-\text{corr}^2} \times \text{z}\right),$$

where

mh	= Mean HT,
sh	= Std Dev HT,
corr	= Correlation coefficient (between log BW and HT),
w	= Simulated log BW,
mw	= Mean Log BW,
sw	= Std Dev Log BW, and
z	= Simulated and truncated standard normal variate.

The z-score "z" is randomly generated from a standard normal distribution. Analogously to the truncation of the BW distribution, **z should be resampled if its absolute value is greater than $z_{0.99}$.**

Table 4. Unsmoothed and Smoothed Parameter Values

		UNSMOOTHED					SMOOTHED					BOUNDS FOR LOG BW	
Age	Gender	Mean HT	Mean Log BW	Std Dev HT	Std Dev Log BW	Correl-ation	Mean HT	Mean Log BW	Std Dev HT	Std Dev Log BW	Correl-ation	Lower Bound	Upper Bound
0	1	66.348	2.024	6.538	0.261	0.934	71.149	2.189	4.541	0.144	0.827	1.855	2.524
1	1	81.551	2.429	4.495	0.126	0.804	79.700	2.362	4.830	0.156	0.817	1.999	2.725
2	1	91.720	2.640	4.508	0.126	0.751	88.191	2.533	5.117	0.168	0.807	2.141	2.924
3	1	98.932	2.773	4.763	0.139	0.742	96.564	2.701	5.399	0.180	0.795	2.283	3.120
4	1	106.749	2.915	4.795	0.157	0.755	104.761	2.867	5.673	0.191	0.783	2.421	3.312
5	1	114.047	3.045	5.750	0.172	0.741	112.722	3.027	5.937	0.202	0.770	2.557	3.498
6	1	119.584	3.149	5.647	0.191	0.758	120.388	3.182	6.188	0.212	0.755	2.689	3.676
7	1	126.274	3.298	6.172	0.234	0.706	127.701	3.330	6.424	0.221	0.738	2.816	3.845
8	1	131.387	3.407	6.487	0.241	0.768	134.601	3.470	6.642	0.229	0.719	2.937	4.004
9	1	137.145	3.513	6.989	0.249	0.721	141.030	3.601	6.840	0.236	0.698	3.052	4.150
10	1	142.600	3.656	6.965	0.263	0.685	146.928	3.721	7.014	0.241	0.673	3.160	4.283
11	1	150.274	3.832	8.441	0.280	0.697	152.251	3.831	7.164	0.245	0.646	3.260	4.401
12	1	155.594	3.894	7.455	0.252	0.671	157.006	3.929	7.290	0.248	0.616	3.352	4.505
13	1	163.822	4.060	8.320	0.256	0.563	161.217	4.016	7.393	0.249	0.585	3.437	4.596
14	1	168.833	4.143	7.825	0.259	0.585	164.906	4.094	7.474	0.250	0.553	3.514	4.675
15	1	173.395	4.234	7.224	0.255	0.485	168.094	4.162	7.535	0.249	0.521	3.583	4.741
16	1	174.662	4.289	6.608	0.226	0.430	170.804	4.222	7.578	0.248	0.491	3.646	4.798
17	1	175.483	4.317	8.067	0.235	0.416	173.059	4.272	7.604	0.246	0.462	3.701	4.844
18	1	175.871	4.363	7.309	0.248	0.451	174.881	4.315	7.613	0.243	0.437	3.749	4.882
19	1	176.655	4.350	7.524	0.229	0.312	176.292	4.350	7.608	0.241	0.415	3.790	4.911
20	1	175.034	4.341	7.566	0.246	0.504	177.314	4.379	7.590	0.238	0.398	3.824	4.933
21	1	176.763	4.342	8.403	0.231	0.426	177.974	4.401	7.561	0.236	0.385	3.852	4.950
22	1	176.195	4.368	6.516	0.224	0.299	178.320	4.417	7.523	0.233	0.378	3.874	4.960
23	1	174.777	4.418	8.261	0.269	0.423	178.401	4.429	7.481	0.231	0.375	3.891	4.967
24	1	176.734	4.396	7.498	0.192	0.391	178.270	4.437	7.437	0.229	0.375	3.904	4.969
25	1	176.400	4.422	6.713	0.248	0.388	177.977	4.441	7.395	0.227	0.377	3.914	4.969
26	1	176.482	4.437	6.841	0.213	0.396	177.575	4.444	7.359	0.225	0.382	3.921	4.967
27	1	176.625	4.433	6.835	0.207	0.515	177.113	4.445	7.333	0.223	0.388	3.926	4.964
28	1	177.668	4.459	7.591	0.230	0.337	176.643	4.446	7.319	0.222	0.394	3.930	4.961
29	1	176.629	4.467	7.984	0.215	0.174	176.217	4.446	7.322	0.220	0.401	3.934	4.959
30	1	177.154	4.458	7.644	0.233	0.597	175.885	4.449	7.344	0.219	0.407	3.939	4.958
31	1	176.424	4.465	6.393	0.211	0.298	175.688	4.452	7.388	0.218	0.411	3.946	4.959

		UNSMOOTHED					SMOOTHED					BOUNDS FOR LOG BW	
32	1	176.506	4.486	8.069	0.230	0.482	175.614	4.458	7.450	0.217	0.414	3.953	4.963
33	1	177.685	4.500	7.686	0.208	0.528	175.643	4.465	7.523	0.216	0.416	3.962	4.967
34	1	176.909	4.451	7.629	0.194	0.292	175.752	4.472	7.603	0.215	0.418	3.971	4.973
35	1	175.465	4.461	8.162	0.228	0.279	175.920	4.480	7.683	0.215	0.418	3.980	4.979
36	1	175.886	4.498	7.555	0.240	0.519	176.124	4.487	7.757	0.214	0.417	3.990	4.985
37	1	176.134	4.512	7.465	0.204	0.434	176.344	4.495	7.821	0.213	0.416	3.999	4.990
38	1	176.737	4.514	7.627	0.210	0.453	176.556	4.501	7.867	0.212	0.415	4.008	4.994
39	1	176.688	4.483	8.195	0.219	0.373	176.740	4.506	7.891	0.211	0.413	4.015	4.996
40	1	177.188	4.504	8.246	0.212	0.546	176.874	4.509	7.886	0.209	0.411	4.022	4.996
41	1	177.129	4.522	8.370	0.218	0.357	176.941	4.510	7.850	0.208	0.410	4.027	4.993
42	1	175.377	4.491	7.477	0.189	0.339	176.945	4.509	7.786	0.206	0.408	4.030	4.988
43	1	177.690	4.510	7.330	0.192	0.367	176.899	4.507	7.701	0.204	0.407	4.033	4.982
44	1	176.112	4.512	7.903	0.202	0.470	176.813	4.504	7.602	0.202	0.406	4.034	4.974
45	1	174.981	4.463	7.396	0.182	0.453	176.697	4.500	7.496	0.200	0.405	4.034	4.966
46	1	176.634	4.485	6.562	0.200	0.227	176.561	4.495	7.389	0.199	0.405	4.033	4.958
47	1	175.600	4.493	6.753	0.180	0.405	176.417	4.491	7.288	0.198	0.406	4.031	4.951
48	1	176.122	4.482	7.434	0.221	0.357	176.276	4.487	7.199	0.197	0.407	4.029	4.945
49	1	177.033	4.488	6.807	0.208	0.496	176.147	4.483	7.130	0.197	0.409	4.025	4.941
50	1	176.496	4.524	7.690	0.199	0.590	176.042	4.481	7.087	0.197	0.412	4.022	4.940
51	1	174.912	4.432	7.901	0.221	0.534	175.968	4.480	7.074	0.199	0.416	4.018	4.942
52	1	176.530	4.501	6.804	0.205	0.338	175.922	4.480	7.089	0.200	0.421	4.013	4.946
53	1	176.744	4.479	7.201	0.215	0.510	175.897	4.480	7.126	0.203	0.427	4.009	4.952
54	1	176.288	4.521	7.453	0.204	0.441	175.887	4.482	7.180	0.205	0.433	4.004	4.959
55	1	175.405	4.483	6.225	0.203	0.363	175.885	4.484	7.246	0.208	0.439	4.000	4.967
56	1	176.729	4.481	7.468	0.192	0.292	175.885	4.486	7.319	0.211	0.444	3.996	4.976
57	1	175.733	4.474	8.368	0.190	0.437	175.880	4.488	7.393	0.213	0.449	3.992	4.984
58	1	176.871	4.474	8.038	0.195	0.324	175.865	4.489	7.463	0.216	0.454	3.988	4.991
59	1	176.603	4.491	6.358	0.217	0.472	175.831	4.490	7.524	0.217	0.457	3.985	4.996
60	1	175.322	4.488	7.743	0.206	0.380	175.774	4.491	7.571	0.219	0.460	3.982	4.999
61	1	175.231	4.506	7.553	0.233	0.387	175.688	4.490	7.600	0.219	0.461	3.980	5.000
62	1	174.979	4.487	7.231	0.192	0.475	175.574	4.488	7.612	0.219	0.460	3.979	4.998
63	1	177.680	4.486	8.229	0.234	0.520	175.436	4.486	7.608	0.218	0.459	3.978	4.993
64	1	173.887	4.467	7.268	0.230	0.534	175.277	4.482	7.591	0.217	0.456	3.977	4.987
65	1	175.770	4.493	7.209	0.229	0.372	175.099	4.478	7.561	0.215	0.453	3.978	4.979
66	1	175.376	4.471	8.807	0.210	0.408	174.906	4.474	7.523	0.213	0.448	3.978	4.970

		UNSMOOTHED					SMOOTHED					BOUNDS FOR LOG BW	
67	1	173.978	4.482	6.767	0.207	0.627	174.701	4.469	7.476	0.211	0.444	3.979	4.959
68	1	174.040	4.447	6.660	0.256	0.490	174.487	4.464	7.424	0.208	0.438	3.980	4.948
69	1	173.767	4.461	8.313	0.204	0.510	174.267	4.458	7.368	0.205	0.433	3.981	4.936
70	1	173.764	4.465	6.780	0.190	0.434	174.043	4.453	7.310	0.203	0.427	3.982	4.924
71	1	171.952	4.442	7.098	0.218	0.413	173.819	4.447	7.252	0.200	0.421	3.983	4.912
72	1	173.617	4.427	7.523	0.184	0.527	173.595	4.442	7.193	0.197	0.416	3.984	4.900
73	1	171.815	4.401	7.548	0.189	0.578	173.371	4.436	7.135	0.194	0.410	3.985	4.888
74	1	173.762	4.452	6.224	0.172	0.220	173.148	4.431	7.076	0.191	0.404	3.986	4.875
75	1	172.609	4.418	7.212	0.219	0.503	172.924	4.425	7.018	0.188	0.399	3.987	4.863
76	1	172.734	4.418	6.328	0.173	0.161	172.700	4.420	6.960	0.185	0.393	3.989	4.851
77	1	172.442	4.457	7.440	0.166	0.400	172.476	4.414	6.901	0.183	0.387	3.990	4.839
78	1	174.156	4.437	7.499	0.199	0.524	172.252	4.409	6.843	0.180	0.381	3.991	4.827
79	1	172.635	4.361	6.417	0.160	0.195	172.028	4.403	6.785	0.177	0.376	3.992	4.814
80	1						171.804	4.398	6.726	0.174	0.370	3.993	4.802
81	1						171.580	4.392	6.668	0.171	0.364	3.994	4.790
82	1						171.357	4.387	6.610	0.168	0.359	3.995	4.778
83	1						171.133	4.381	6.551	0.165	0.353	3.996	4.766
84	1						170.909	4.376	6.493	0.162	0.347	3.998	4.754
85	1						170.685	4.370	6.434	0.160	0.341	3.999	4.741
86	1						170.461	4.365	6.376	0.157	0.336	4.000	4.729
87	1						170.237	4.359	6.318	0.154	0.330	4.001	4.717
88	1						170.013	4.353	6.259	0.151	0.324	4.002	4.705
89	1						169.789	4.348	6.201	0.148	0.319	4.003	4.693
90	1						169.565	4.342	6.143	0.145	0.313	4.004	4.680
91	1						169.342	4.337	6.084	0.142	0.307	4.006	4.668
92	1						169.118	4.331	6.026	0.140	0.301	4.007	4.656
93	1						168.894	4.326	5.968	0.137	0.296	4.008	4.644
94	1						168.670	4.320	5.909	0.134	0.290	4.009	4.632
95	1						168.446	4.315	5.851	0.131	0.284	4.010	4.620
96	1						168.222	4.309	5.792	0.128	0.279	4.011	4.607
97	1						167.998	4.304	5.734	0.125	0.273	4.012	4.595
98	1						167.774	4.298	5.676	0.122	0.267	4.013	4.583
99	1						167.550	4.293	5.617	0.120	0.262	4.015	4.571
100	1						167.327	4.287	5.559	0.117	0.256	4.016	4.559
0	2	64.997	1.963	6.275	0.270	0.933	68.702	2.113	4.597	0.164	0.848	1.731	2.495

		UNSMOOTHED					SMOOTHED					BOUNDS FOR LOG BW	
1	2	80.615	2.394	4.947	0.152	0.789	77.867	2.301	4.849	0.173	0.831	1.898	2.705
2	2	89.528	2.573	4.204	0.133	0.765	86.943	2.488	5.097	0.183	0.813	2.063	2.912
3	2	98.281	2.739	4.248	0.145	0.733	95.842	2.671	5.339	0.192	0.794	2.225	3.116
4	2	105.404	2.879	4.857	0.168	0.761	104.476	2.849	5.571	0.200	0.775	2.383	3.314
5	2	112.415	3.012	5.787	0.181	0.744	112.756	3.020	5.790	0.208	0.754	2.535	3.505
6	2	118.957	3.147	5.654	0.205	0.734	120.594	3.184	5.993	0.216	0.731	2.681	3.686
7	2	124.658	3.261	5.843	0.238	0.753	127.901	3.337	6.177	0.223	0.706	2.818	3.857
8	2	131.786	3.433	6.950	0.253	0.720	134.589	3.479	6.338	0.230	0.679	2.944	4.014
9	2	137.722	3.566	6.500	0.264	0.676	140.569	3.608	6.474	0.236	0.649	3.060	4.156
10	2	144.426	3.681	7.298	0.256	0.729	145.754	3.722	6.581	0.240	0.616	3.163	4.281
11	2	150.574	3.818	7.670	0.278	0.606	150.083	3.820	6.657	0.244	0.580	3.252	4.389
12	2	156.583	3.963	7.295	0.273	0.558	153.611	3.904	6.705	0.247	0.541	3.328	4.479
13	2	158.923	4.000	6.149	0.242	0.391	156.424	3.974	6.730	0.250	0.501	3.393	4.555
14	2	160.849	4.069	6.429	0.209	0.344	158.606	4.032	6.737	0.251	0.460	3.447	4.617
15	2	161.704	4.126	6.674	0.240	0.461	160.241	4.079	6.728	0.253	0.420	3.491	4.667
16	2	162.002	4.140	6.219	0.251	0.364	161.413	4.118	6.710	0.254	0.382	3.528	4.708
17	2	162.805	4.151	6.661	0.244	0.359	162.208	4.149	6.687	0.254	0.347	3.558	4.740
18	2	162.208	4.177	6.344	0.263	0.228	162.709	4.174	6.662	0.255	0.315	3.582	4.766
19	2	163.320	4.193	6.174	0.266	0.227	163.000	4.195	6.640	0.255	0.288	3.602	4.788
20	2	163.411	4.175	7.485	0.250	0.294	163.167	4.213	6.626	0.255	0.266	3.619	4.807
21	2	161.858	4.194	6.643	0.253	0.397	163.281	4.229	6.624	0.256	0.252	3.634	4.825
22	2	162.038	4.263	6.058	0.257	0.086	163.358	4.244	6.632	0.257	0.243	3.647	4.842
23	2	161.916	4.269	7.447	0.258	0.294	163.405	4.258	6.649	0.258	0.239	3.658	4.857
24	2	162.774	4.270	7.195	0.259	0.236	163.425	4.270	6.675	0.259	0.240	3.667	4.872
25	2	162.763	4.235	6.405	0.261	0.288	163.423	4.280	6.707	0.260	0.244	3.676	4.885
26	2	163.198	4.278	6.312	0.268	0.325	163.404	4.289	6.744	0.261	0.251	3.683	4.896
27	2	163.593	4.300	7.471	0.272	0.356	163.372	4.297	6.786	0.262	0.260	3.689	4.906
28	2	163.380	4.293	6.569	0.274	0.354	163.332	4.304	6.829	0.262	0.271	3.694	4.914
29	2	162.909	4.305	5.527	0.234	0.269	163.288	4.309	6.874	0.263	0.281	3.698	4.920
30	2	163.515	4.318	7.695	0.262	0.269	163.246	4.314	6.919	0.263	0.292	3.702	4.925
31	2	164.013	4.316	6.712	0.264	0.212	163.208	4.316	6.962	0.263	0.301	3.705	4.927
32	2	163.674	4.331	7.194	0.289	0.248	163.176	4.318	7.002	0.262	0.309	3.708	4.928
33	2	163.856	4.353	6.710	0.243	0.269	163.148	4.319	7.039	0.262	0.315	3.711	4.928
34	2	163.344	4.341	7.496	0.278	0.283	163.124	4.319	7.072	0.261	0.320	3.713	4.926
35	2	163.531	4.309	6.544	0.221	0.200	163.103	4.319	7.100	0.260	0.323	3.715	4.923

		UNSMOOTHED					SMOOTHED					BOUNDS FOR LOG BW	
36	2	163.211	4.304	7.656	0.259	0.362	163.085	4.318	7.122	0.259	0.325	3.717	4.920
37	2	164.099	4.333	6.902	0.276	0.391	163.070	4.317	7.137	0.257	0.324	3.719	4.916
38	2	162.956	4.354	7.860	0.236	0.328	163.056	4.316	7.145	0.256	0.322	3.720	4.913
39	2	162.702	4.305	7.675	0.251	0.396	163.043	4.316	7.144	0.255	0.318	3.722	4.909
40	2	162.678	4.303	7.397	0.279	0.302	163.031	4.315	7.134	0.254	0.311	3.724	4.906
41	2	161.638	4.301	6.643	0.260	0.367	163.018	4.315	7.114	0.253	0.302	3.726	4.904
42	2	163.154	4.298	7.131	0.264	0.300	163.004	4.315	7.085	0.252	0.291	3.729	4.902
43	2	162.756	4.311	6.773	0.243	0.233	162.987	4.316	7.050	0.251	0.280	3.731	4.901
44	2	162.821	4.290	6.921	0.260	0.301	162.965	4.317	7.010	0.251	0.267	3.734	4.901
45	2	162.737	4.290	5.720	0.238	0.240	162.937	4.319	6.967	0.250	0.255	3.736	4.901
46	2	162.146	4.348	7.539	0.240	0.245	162.902	4.320	6.923	0.250	0.243	3.739	4.901
47	2	163.495	4.360	7.326	0.249	0.254	162.858	4.322	6.879	0.249	0.233	3.742	4.902
48	2	163.566	4.324	6.311	0.260	0.042	162.803	4.324	6.837	0.249	0.224	3.745	4.903
49	2	162.858	4.338	6.338	0.240	0.262	162.737	4.326	6.800	0.249	0.218	3.748	4.904
50	2	162.498	4.345	6.919	0.221	0.248	162.657	4.328	6.768	0.248	0.215	3.750	4.906
51	2	162.610	4.320	5.990	0.244	0.167	162.563	4.330	6.743	0.248	0.215	3.753	4.907
52	2	161.654	4.267	7.879	0.252	0.347	162.456	4.332	6.725	0.248	0.218	3.756	4.908
53	2	163.379	4.325	6.657	0.243	0.260	162.336	4.334	6.713	0.247	0.223	3.758	4.909
54	2	162.049	4.368	7.027	0.262	0.235	162.207	4.335	6.706	0.247	0.231	3.761	4.910
55	2	162.694	4.364	6.633	0.270	0.178	162.068	4.337	6.703	0.247	0.240	3.763	4.911
56	2	162.638	4.332	6.787	0.245	0.115	161.922	4.338	6.703	0.246	0.249	3.764	4.911
57	2	160.512	4.320	7.084	0.225	0.301	161.770	4.338	6.705	0.246	0.259	3.766	4.910
58	2	160.963	4.372	7.017	0.306	0.287	161.613	4.338	6.708	0.245	0.269	3.767	4.909
59	2	160.849	4.305	6.991	0.223	0.266	161.454	4.338	6.712	0.245	0.278	3.768	4.907
60	2	161.262	4.349	6.422	0.276	0.414	161.293	4.337	6.716	0.244	0.286	3.769	4.905
61	2	163.010	4.374	7.148	0.215	0.380	161.131	4.335	6.718	0.243	0.292	3.769	4.901
62	2	160.395	4.373	6.512	0.216	0.266	160.970	4.333	6.719	0.242	0.296	3.769	4.897
63	2	161.629	4.282	6.589	0.229	0.310	160.808	4.330	6.719	0.241	0.299	3.769	4.892
64	2	160.269	4.333	6.028	0.243	0.248	160.647	4.327	6.718	0.240	0.300	3.768	4.886
65	2	161.070	4.284	6.539	0.219	0.240	160.485	4.324	6.716	0.239	0.301	3.768	4.880
66	2	159.425	4.320	5.689	0.207	0.331	160.324	4.320	6.713	0.238	0.300	3.766	4.873
67	2	160.241	4.318	6.903	0.259	0.351	160.163	4.316	6.710	0.237	0.299	3.765	4.866
68	2	158.931	4.237	7.056	0.232	0.300	160.001	4.311	6.707	0.235	0.297	3.764	4.858
69	2	159.863	4.288	6.687	0.237	0.287	159.839	4.307	6.703	0.234	0.295	3.763	4.851
70	2	160.263	4.361	6.986	0.243	0.257	159.678	4.302	6.699	0.233	0.292	3.761	4.843

		UNSMOOTHED					SMOOTHED					BOUNDS FOR LOG BW	
71	2	159.678	4.318	7.340	0.259	0.275	159.516	4.298	6.695	0.231	0.290	3.760	4.836
72	2	158.699	4.295	6.225	0.254	0.262	159.355	4.293	6.691	0.230	0.287	3.758	4.828
73	2	159.618	4.307	7.187	0.234	0.302	159.193	4.288	6.687	0.229	0.285	3.757	4.820
74	2	159.042	4.280	6.425	0.216	0.237	159.032	4.284	6.683	0.227	0.282	3.755	4.812
75	2	158.332	4.276	7.461	0.222	0.083	158.870	4.279	6.679	0.226	0.280	3.754	4.805
76	2	159.769	4.324	5.740	0.209	0.297	158.709	4.275	6.675	0.225	0.277	3.752	4.797
77	2	158.186	4.270	5.841	0.238	0.248	158.547	4.270	6.671	0.223	0.275	3.751	4.789
78	2	158.001	4.258	7.098	0.216	0.292	158.386	4.266	6.667	0.222	0.272	3.750	4.782
79	2	158.586	4.224	7.097	0.208	0.461	158.224	4.261	6.663	0.220	0.270	3.748	4.774
80	2						158.063	4.257	6.659	0.219	0.267	3.747	4.766
81	2						157.901	4.252	6.655	0.218	0.265	3.745	4.759
82	2						157.740	4.247	6.651	0.216	0.262	3.744	4.751
83	2						157.578	4.243	6.648	0.215	0.260	3.742	4.743
84	2						157.417	4.238	6.644	0.214	0.257	3.741	4.736
85	2						157.255	4.234	6.640	0.212	0.255	3.739	4.728
86	2						157.094	4.229	6.636	0.211	0.252	3.738	4.720
87	2						156.932	4.225	6.632	0.210	0.250	3.737	4.712
88	2						156.771	4.220	6.628	0.208	0.247	3.735	4.705
89	2						156.609	4.215	6.624	0.207	0.245	3.734	4.697
90	2						156.448	4.211	6.620	0.206	0.242	3.732	4.689
91	2						156.286	4.206	6.616	0.204	0.240	3.731	4.682
92	2						156.125	4.202	6.612	0.203	0.237	3.729	4.674
93	2						155.963	4.197	6.608	0.202	0.235	3.728	4.666
94	2						155.802	4.193	6.604	0.200	0.232	3.727	4.659
95	2						155.640	4.188	6.600	0.199	0.230	3.725	4.651
96	2						155.479	4.183	6.596	0.198	0.227	3.724	4.643
97	2						155.317	4.179	6.592	0.196	0.225	3.722	4.636
98	2						155.156	4.174	6.588	0.195	0.222	3.721	4.628
99	2						154.994	4.170	6.584	0.194	0.220	3.719	4.620
100	2						154.833	4.165	6.580	0.192	0.217	3.718	4.613

5. Comparison between 2009–2014 and 2003–2014

The fitted models for 2009–2014 are contained in Table 4 and in the tab “Parameters” of the accompanying Excel file “means.2009 to 2014.102016.xlsx”. We give unsmoothed and smoothed parameters for each age and gender. Using the same approach, the fitted parameters for 2003–2014 are contained in the tab “Parameters” of the accompanying Excel file “means.2003 to 2014.102016.xlsx”.

The following Table 5 contains a comparison of the parameters between the two sets of years. The differences and percentage differences are relative to the baseline of 2003–2014:

$$\text{Difference} = \text{Value for } 2009\text{–}2014 \text{ - Value for } 2003\text{–}2014$$

$$\text{Percentage Difference} = \text{Difference} \,/\, \text{Value for } 2003\text{–}2014 \times 100$$

The tabulated means and maxima are for each gender across all ages 0–79 years, for both the unsmoothed and smoothed parameters.

The mean differences are between -0.14 and 0.07 across all parameters, so there is only a small trend in the parameters. (Note that the two periods overlap, but any difference between the overlapping periods implies a difference between 2003–2008 and 2009–2014.)

The differences are small for the mean parameters: the maximum unsigned percentage differences are at most 1.7 percent for the unsmoothed mean parameters and at most 0.6 percent for the smoothed mean parameters.

The differences are much higher for the standard deviations and the correlations. For the unsmoothed data, the maximum unsigned percentage difference is 17 percent for the standard deviation of the HT and 69 percent for the correlation. For the smoothed data, the differences are much smaller: the maximum unsigned percentage difference is 5.4 percent for the standard deviation of the HT and 10.7 percent for the correlation.

The mean unsigned percentage difference is at most 13.7 percent across all unsmoothed parameters and at most 3.4 percent across all smoothed parameters.

The lack of a large trend between the two time periods, and the small percentage differences for the smoothed parameters, suggest that it will not make very much difference which set of years is used for the APEX model inputs. **We recommend using the more recent data from 2009–2014**.

Table 5. Differences between Parameters for 2009–2014 and 2003–2014 (Baseline)

	Statistic	Gender	Mean Difference	Mean Percentage Difference	Mean Unsigned Percentage Difference	Maximum Unsigned Percentage Difference
Unsmoothed	Mean HT	1	-0.12	-0.07	0.22	0.86
	Mean HT	2	-0.14	-0.09	0.23	0.67
	Mean Log BW	1	0.00	0.05	0.27	0.91
	Mean Log BW	2	0.01	0.16	0.34	1.65
	Std Dev HT	1	-0.04	-0.57	4.19	17.42
	Std Dev HT	2	0.07	0.96	4.47	10.08
	Std Dev Log BW	1	0.00	0.49	3.82	11.59
	Std Dev Log BW	2	0.00	1.04	4.09	13.49
	Correlation	1	-0.01	-1.67	10.65	51.40
	Correlation	2	0.00	0.22	13.71	68.67
Smoothed	Mean HT	1	-0.12	-0.07	0.12	0.32
	Mean HT	2	-0.14	-0.09	0.12	0.40
	Mean Log BW	1	0.00	0.05	0.08	0.21
	Mean Log BW	2	0.01	0.17	0.19	0.61
	Std Dev HT	1	-0.04	-0.58	1.69	5.41
	Std Dev HT	2	0.07	1.00	1.48	4.42
	Std Dev Log BW	1	0.00	0.50	1.27	2.98
	Std Dev Log BW	2	0.00	1.06	1.20	4.28
	Correlation	1	-0.01	-1.16	2.19	7.00
	Correlation	2	0.00	-1.21	3.37	10.71

6. Effect on HT and WT in APEX using Updated Algorithm

6.1. Description of APEX Runs and Analysis

To summarize the effect of the new algorithm on simulated HT and WT values, we conducted two separate APEX runs: one employing the HT and BW calculations based on the 1999–2004 NHANES data (referred to as the “old method” in this section) and one employing the HT and BW calculation method based on the 2009–2014 NHANES data as proposed in this memorandum (the “new method”). Apart from this difference, the two APEX runs were identical. Both APEX runs employed 100,000 profiles and modeled ages 0–99 years old. This produced a set of 100,000 HT, WT, and body mass index (BMI) values (one of each for each profile).

We analyzed statistics of the HT, WT, and BMI of the profiles generated in APEX for each of 14 age bins. We created the age bins so that they each (except for the oldest bin) contained a roughly equal number of profiles: 5-year bins ages 0–55 years, then single bins for 55–62 years, 62–75 years, and 75–99 years. We present in Figure 6-1 the number of profiles in each age bin.

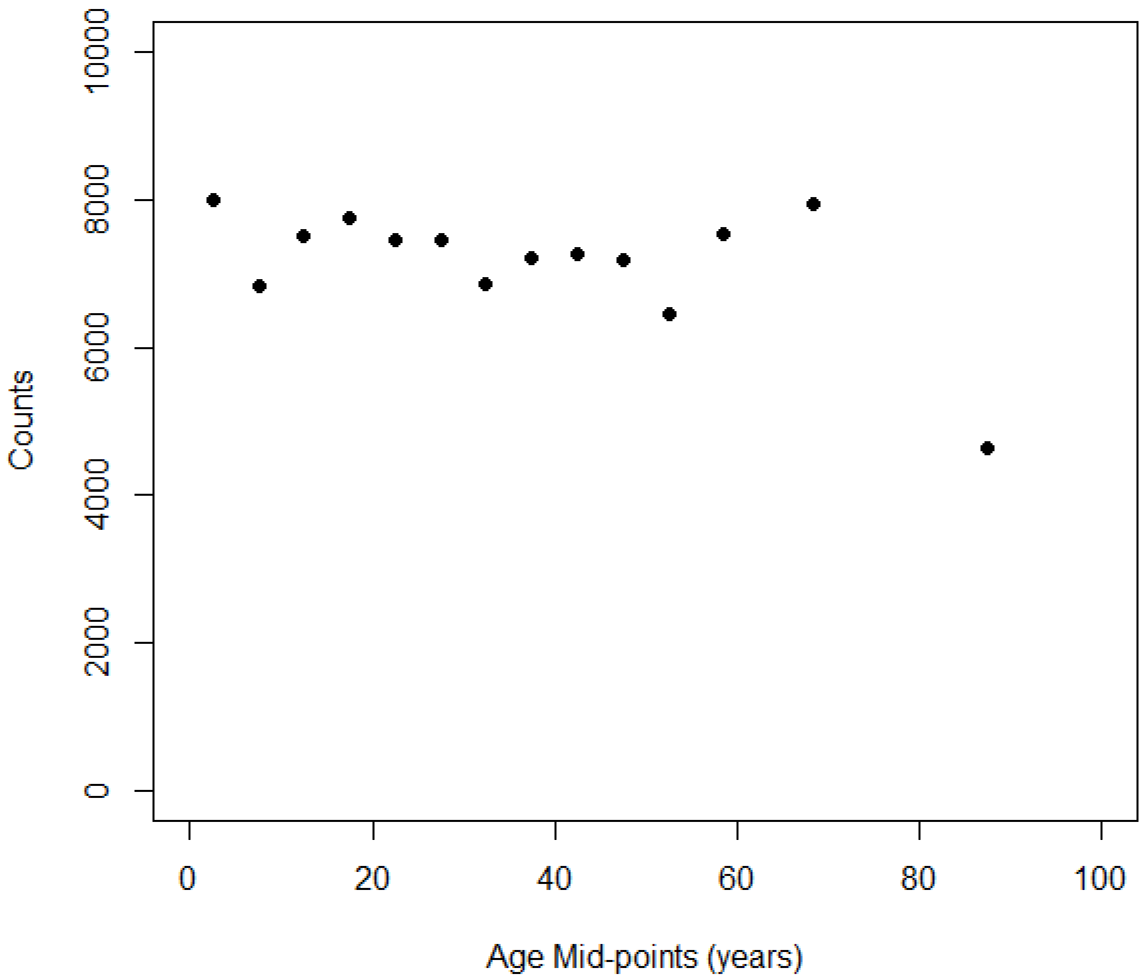


Figure 6-1. Number of Profiles in each Age Bin from APEX Runs (100,000 profiles)

6.2. Comparison of HT, WT, and BMI Results

Table 6 presents a statistical summary and comparison of the HT, WT, and BMI values generated in the two APEX runs employing the old and new methods. These statistics were calculated only on the basis of gender and not on the basis of age bin.

We also compared the outputs of the two methods on the basis of age bin. Figure 6-2 through Figure 6-7 present the mean and standard deviation of HT, WT, and BMI values from the old and new methods in each age bin for the 100,000 profiles generated in APEX.

Table 6. Statistical Summary of HT, WT, and BMI in APEX using Old and New Methods

Variable		Gender	N	Mean	St. Dev	Min	Max	% Difference in Mean
Height (cm)	Old	M	49,473	164.948	25.582	63.058	205.788	-0.108
	New		49,473	164.770	26.038	58.240	205.776	
	Old	F	50,527	154.176	20.525	63.251	187.350	0.126
	New		50,527	154.371	21.230	54.668	190.061	
Weight (kg)	Old	M	49,473	73.943	28.745	3.600	199.198	2.085
	New		49,473	75.484	29.782	6.392	148.412	
	Old	F	50,527	65.056	24.744	3.700	165.998	2.373
	New		50,527	66.600	25.885	5.646	138.102	
BMI (kg/m^2)	Old	M	49,473	25.611	6.374	5.385	59.404	2.075
	New		49,473	26.143	6.637	10.162	54.052	
	Old	F	50,527	26.189	7.440	5.491	63.184	1.824
	New		50,527	26.667	7.690	10.155	61.574	

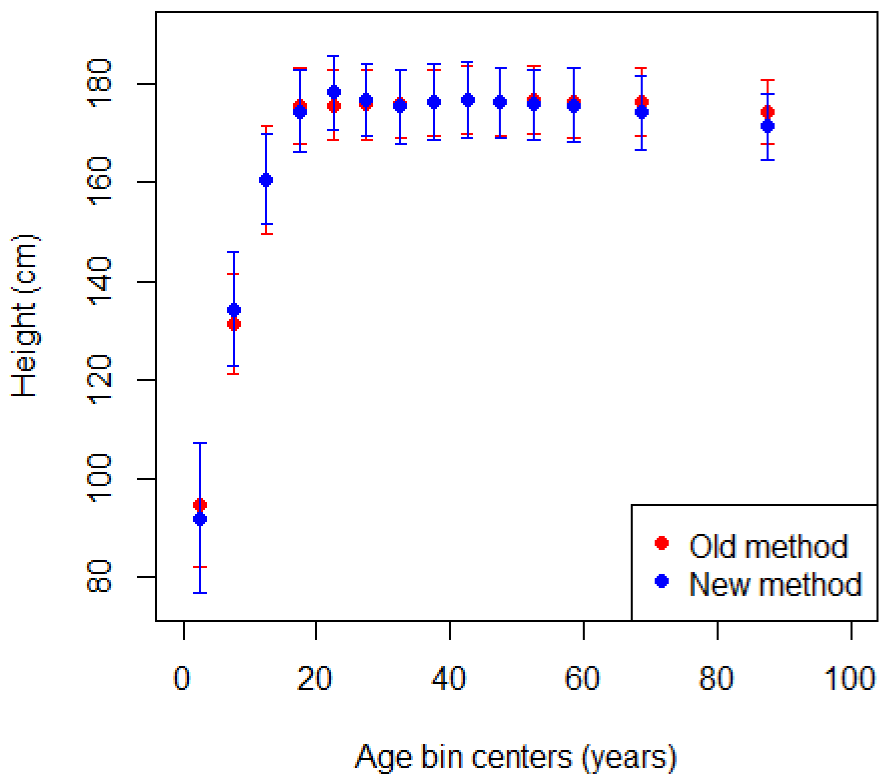


Figure 6-2. Mean ± Standard Deviation of HT for Males

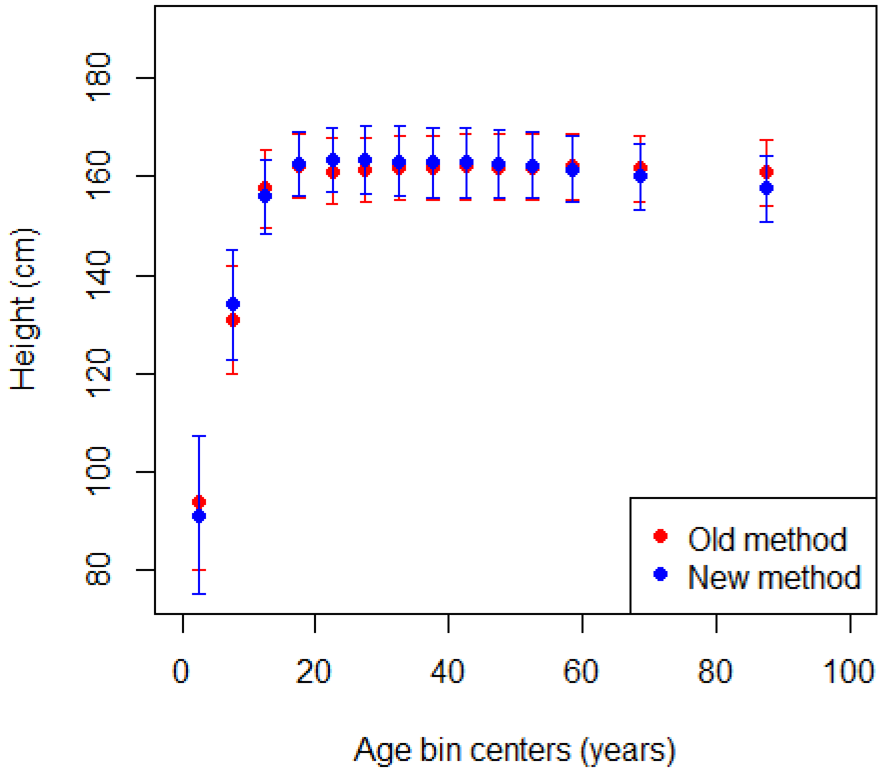


Figure 6-3. Mean ± Standard Deviation of HT for Females

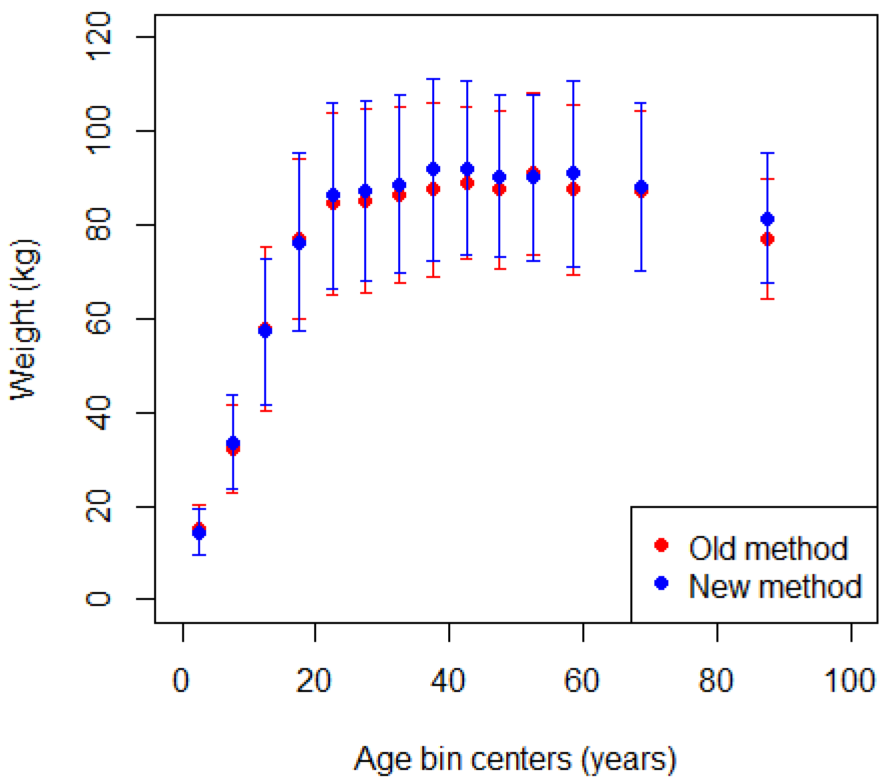


Figure 6-4. Mean ± Standard Deviation of WT for Males

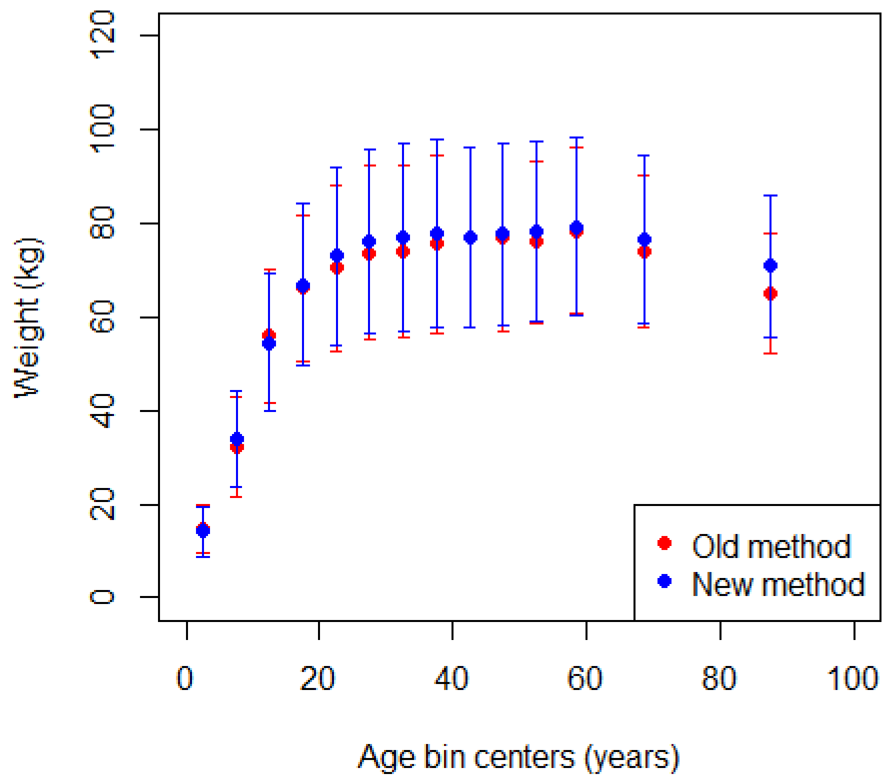


Figure 6-5. Mean ± Standard Deviation of WT for Females

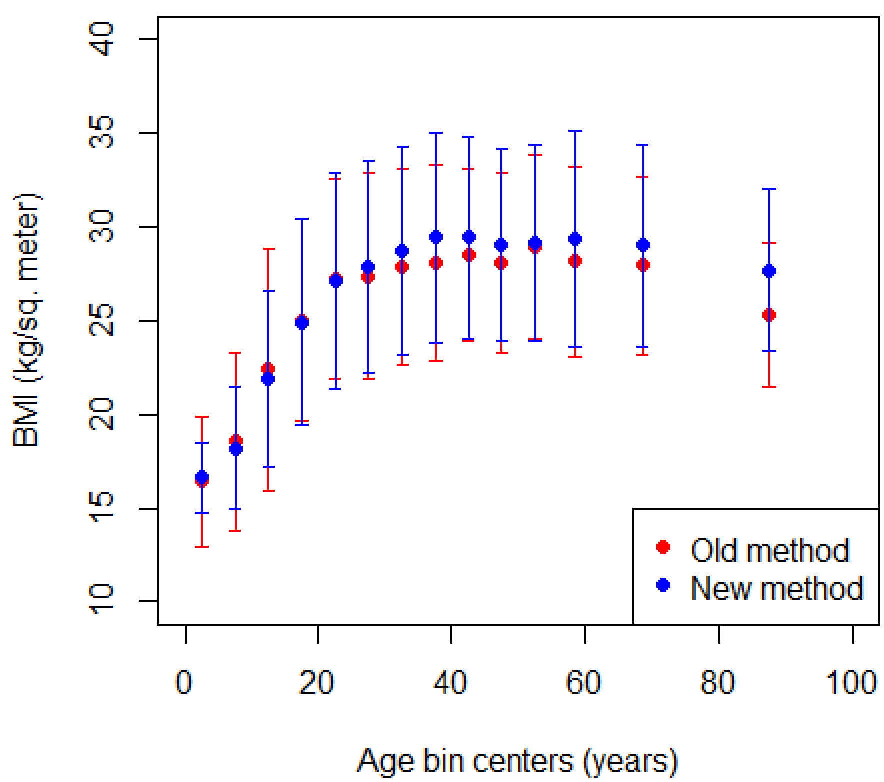


Figure 6-6. Mean ± Standard Deviation of BMI for Males

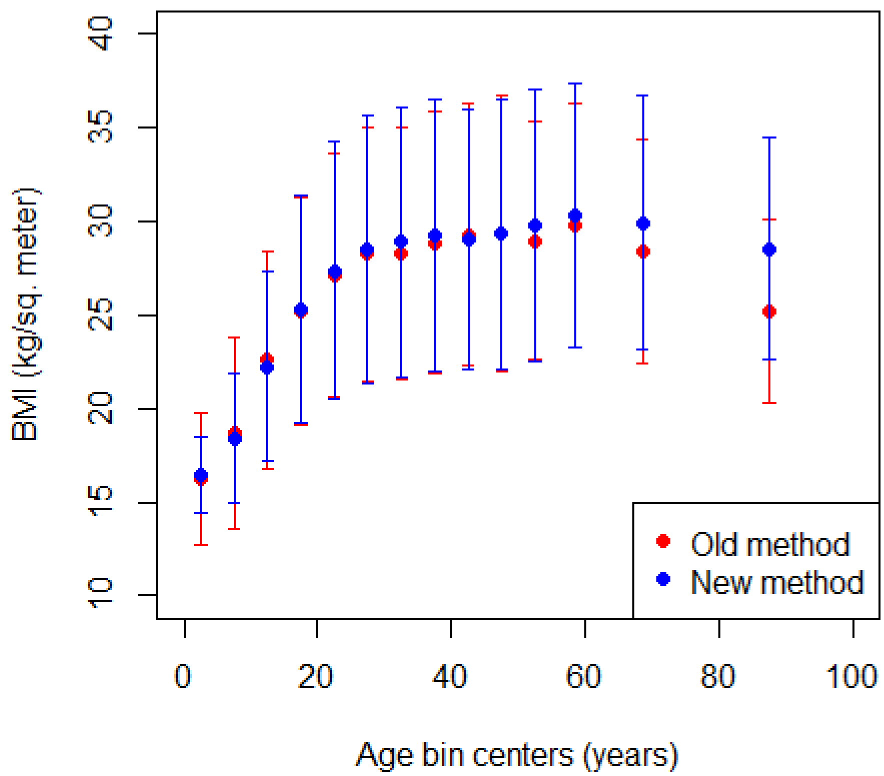


Figure 6-7. Mean ± Standard Deviation of BMI for Females

We made the following observations based on the information presented above comparing the results of the new method to those of the old method:

- When analyzing results irrespective of age bin, the percent differences in the means between the old and new methods for all parameters are small: about 0.1 percent for HT determination (negative for males, positive for females) and about +2.0 percent for the WT and BMI determinations.
- For both males and females, profiles in the youngest age bin (0–5 years) and in the oldest several age bins (from about 55 years and older) are slightly shorter when employing the new method. The old method used in APEX was known to occasionally generate HTs that were too tall for these age groups—for children because HT was not correlated with BW, and for older adults because HT was not correlated with age. This average decrease in HT values reflects the expected change that would occur when including these dependent variables.
- While not consistent across age bins, profiles of both genders are generally heavier using the new method (most apparent with adults, except for males and especially females around ages 40–55 years). This increase can be seen in both the mean values and in the mean ± standard deviation values. This likely reflects trends in WT for the U.S. population (the new method uses newer NHANES data than those of the old method). At the far ends of the simulated WT distribution, the new method estimates higher WT values for the lightest profiles and lower WT values for the heaviest profiles.

- For BMI values, the new method substantially decreased the standard deviation for ages 0–15 years, with generally lower BMI means as well (except in the youngest age group). For adults, there is a general increase in the means and standard deviations of BMI values using the new method, especially for males.
- For a previous assessment, we generated the distribution of BMI values shown in Figure 6-8, from NHANES 2003–2014 data. The distributions of BMI values in these simulations are similar to the NHANES BMI distributions. The majority of BMI values from NHANES are between about 15 and 35 kg/m^2, and the mean BMI values simulated here also fall within that range. BMI values below 15 kg/m^2 and above 40 kg/m^2 are relatively rare in the NHANES data, and the same is true of the BMI values simulated here.

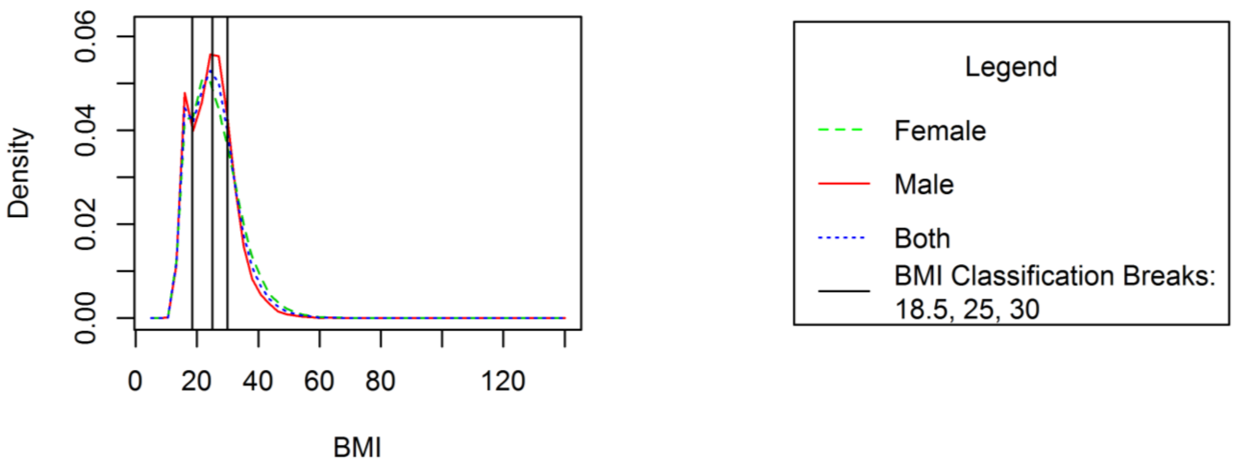
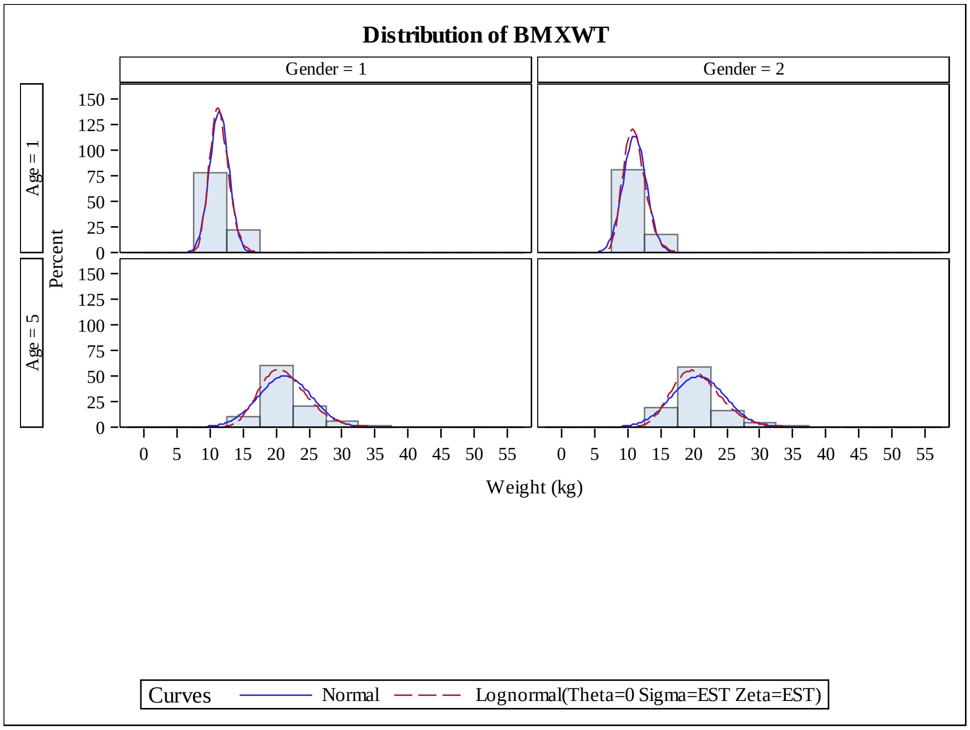
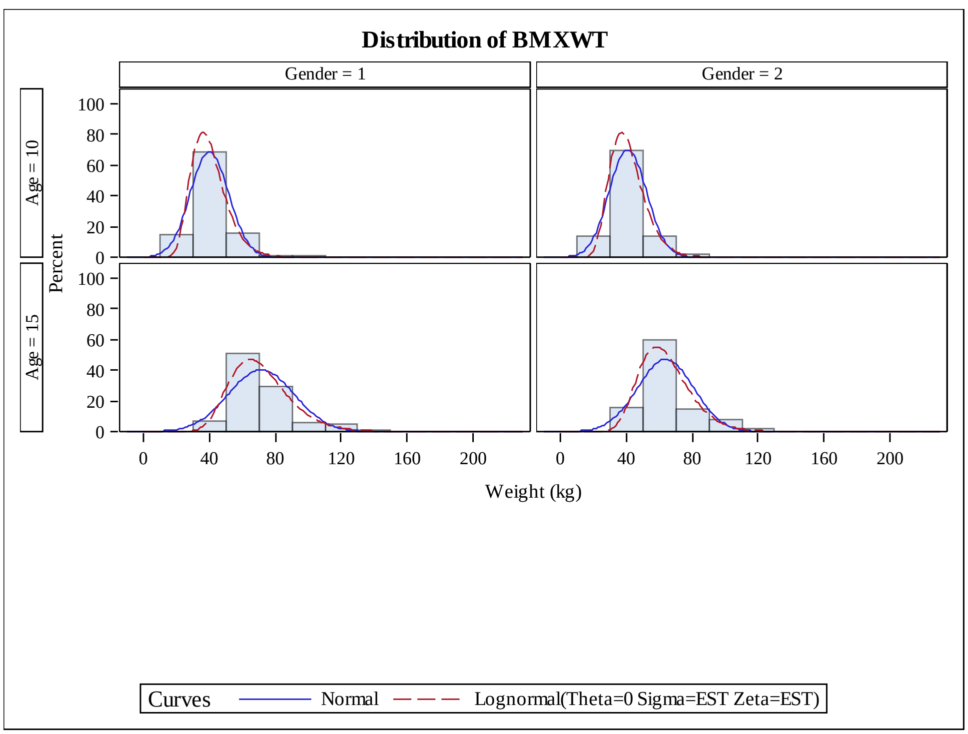
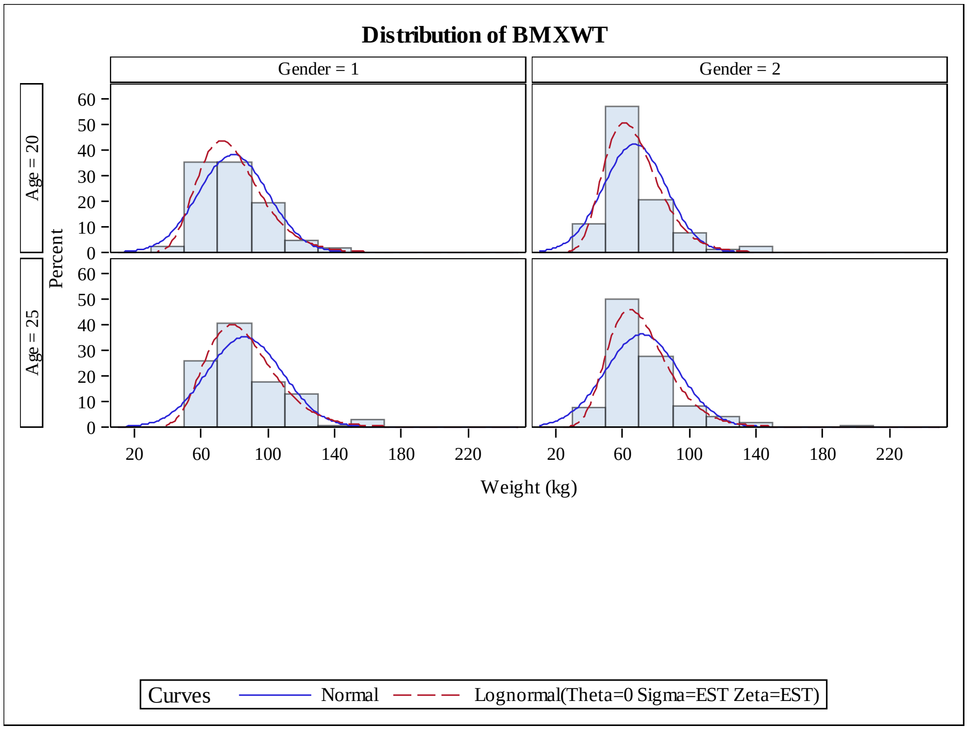


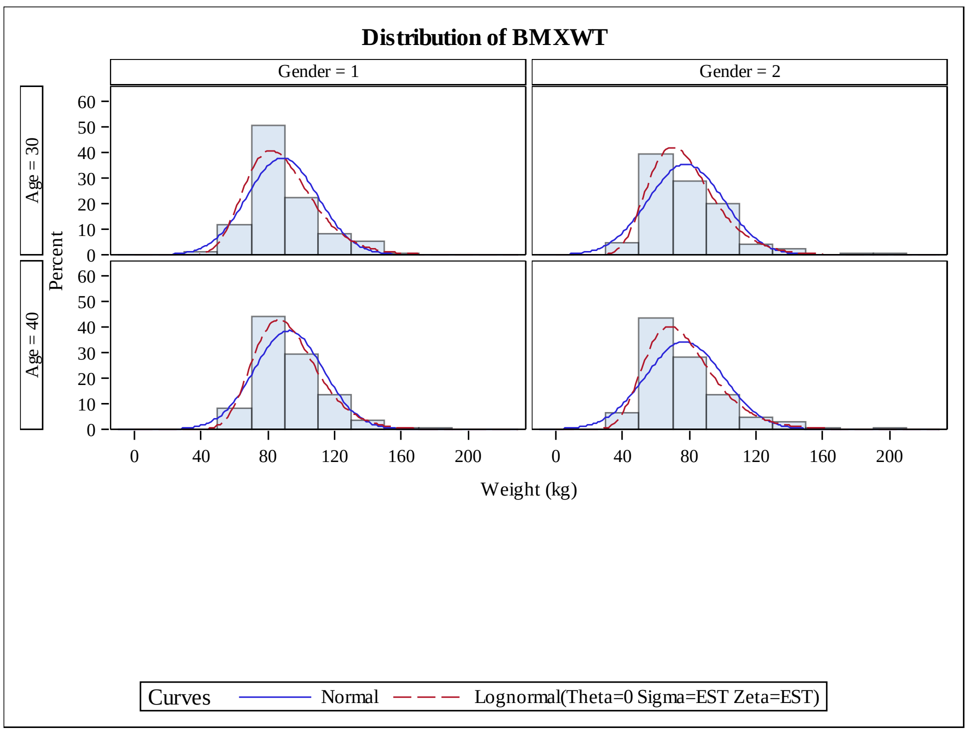
Figure 6-8. Distribution of BMI Values (kg/m^2) from NHANES 2003–2014

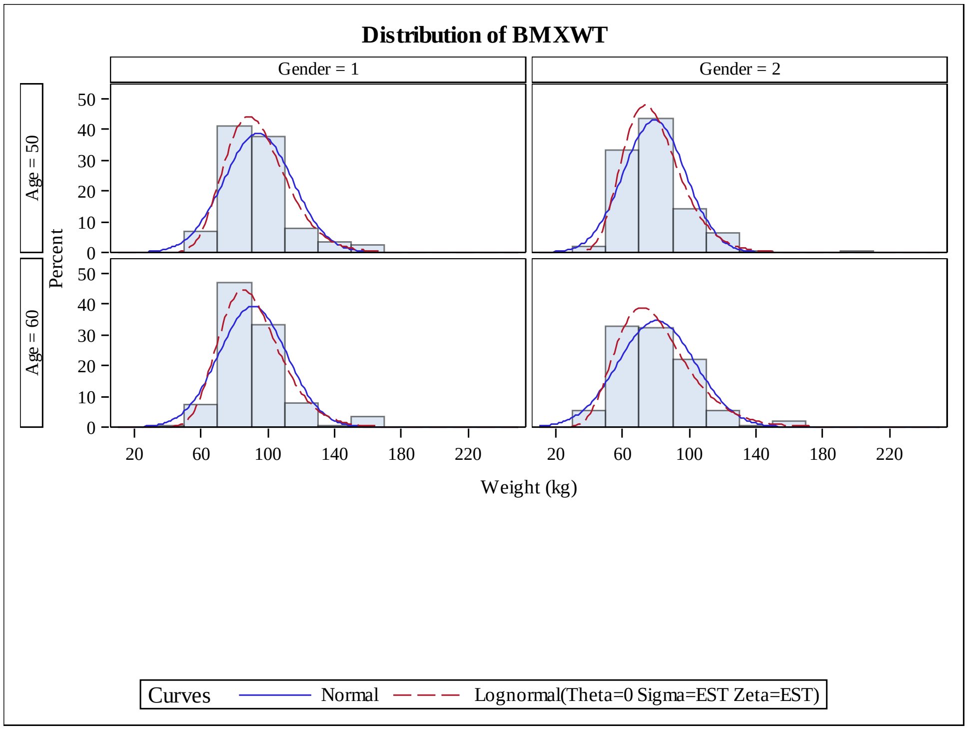
Attachment A. Distributions of Body Weight

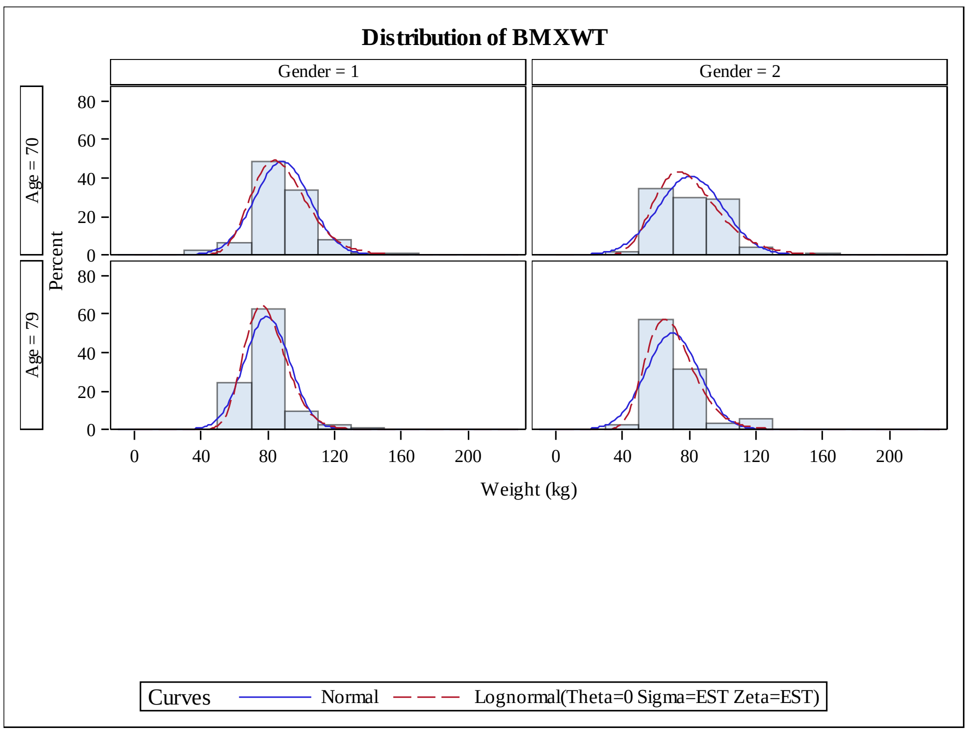




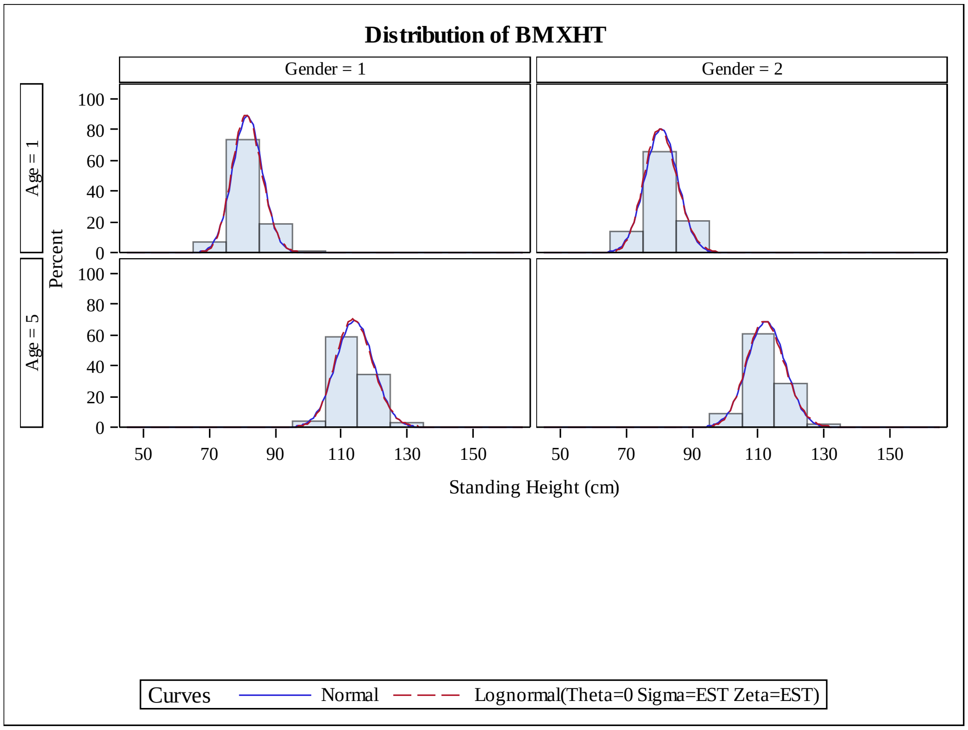

Distribution of BMXWT
Gender = 1
Gender = 2
Age = 20
Age = 25
Percent
0
10
20
30
40
50
60
20
60
100
140
180
220
Weight (kg)
Curves
Normal
Lognormal(Theta=0 Sigma=EST Zeta=EST)

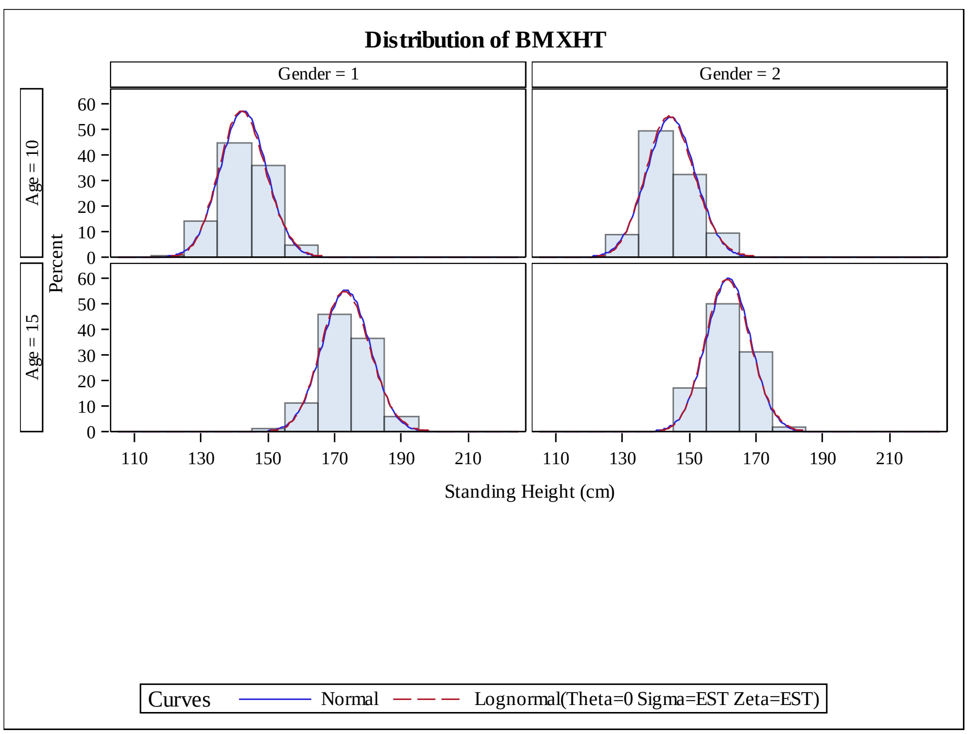

Distribution of BMXWT
Gender = 1
Gender = 2
Age = 30
Age = 40
Percent
0
10
20
30
40
50
60
0
40
80
120
160
200
Weight (kg)
Curves
Normal
Lognormal(Theta=0 Sigma=EST Zeta=EST)

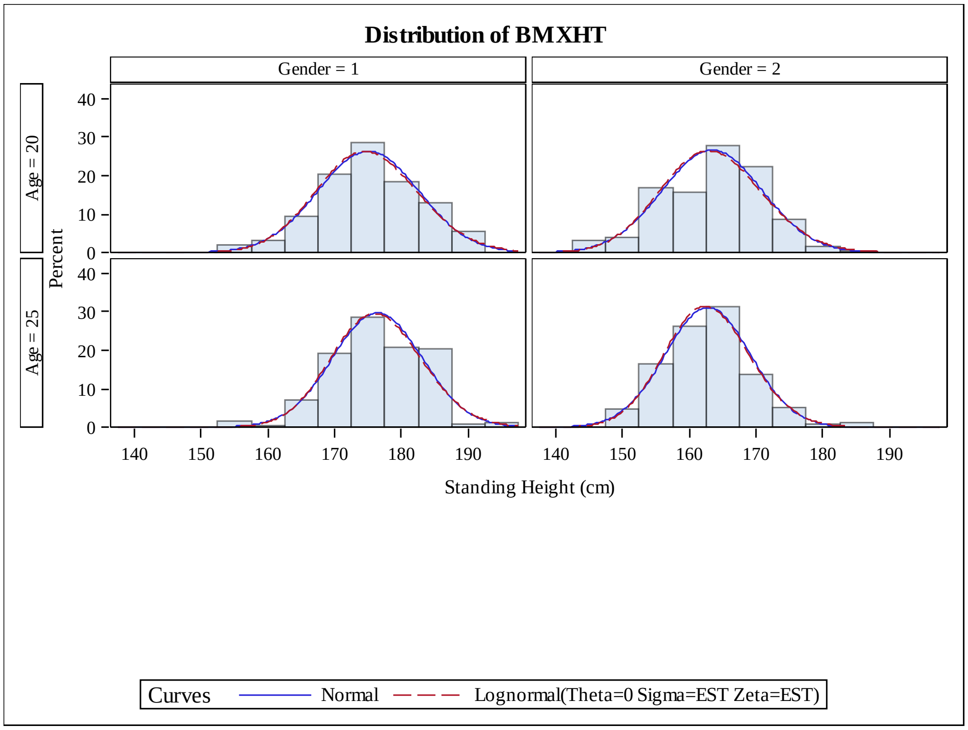

Distribution of BMXWT
Gender = 1
Gender = 2
Age = 50
Age = 60
Percent
Weight (kg)
Curves
Normal
Lognormal(Theta=0 Sigma=EST Zeta=EST)

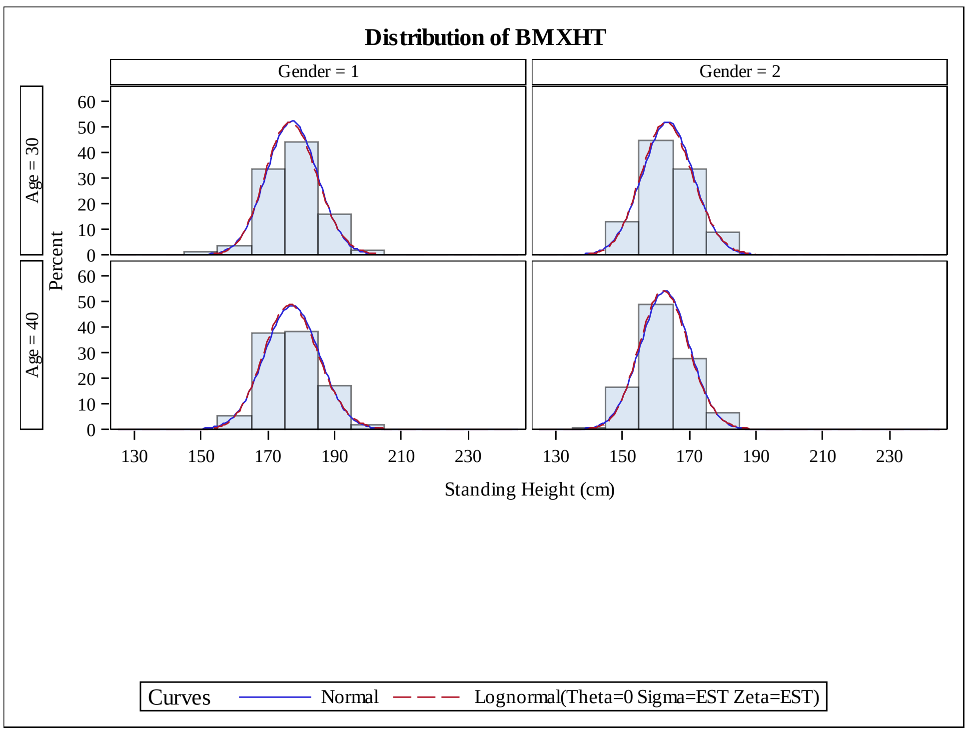

Distribution of BMXWT
Gender = 1
Gender = 2
Age = 70
Age = 79
Percent
Weight (kg)
Curves
Normal
Lognormal(Theta=0 Sigma=EST Zeta=EST)

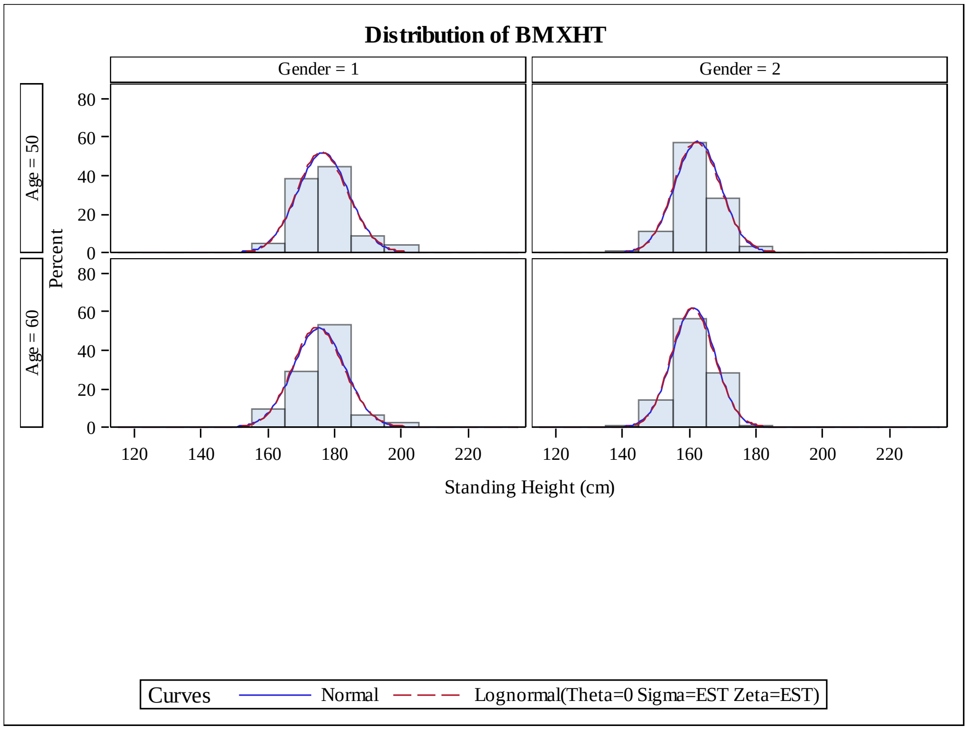
Attachment B. Distributions of Height

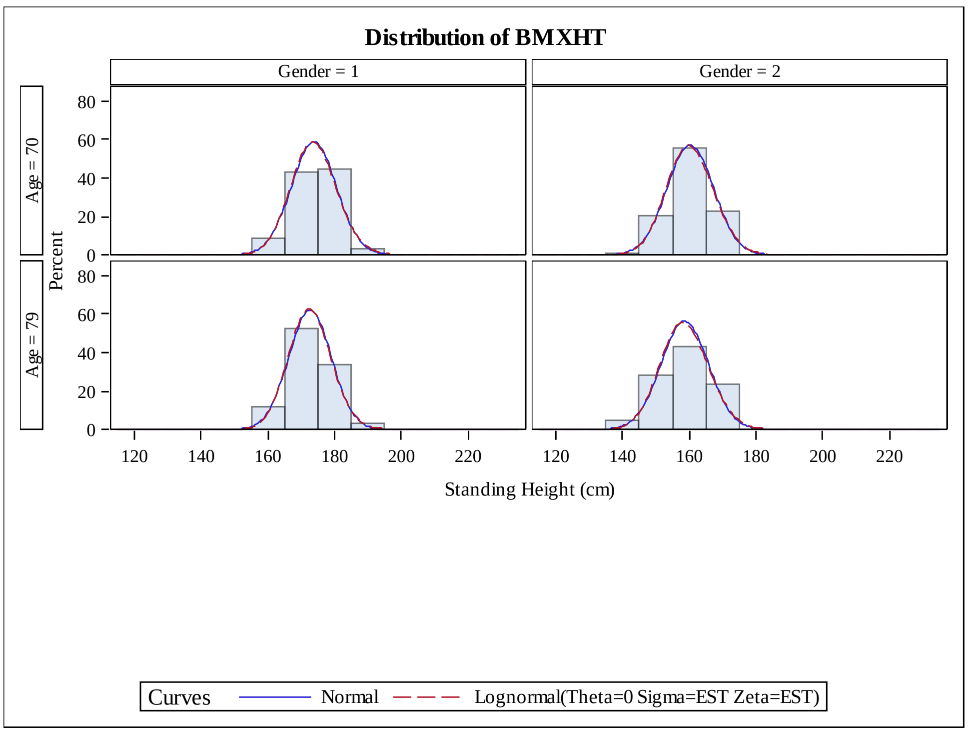




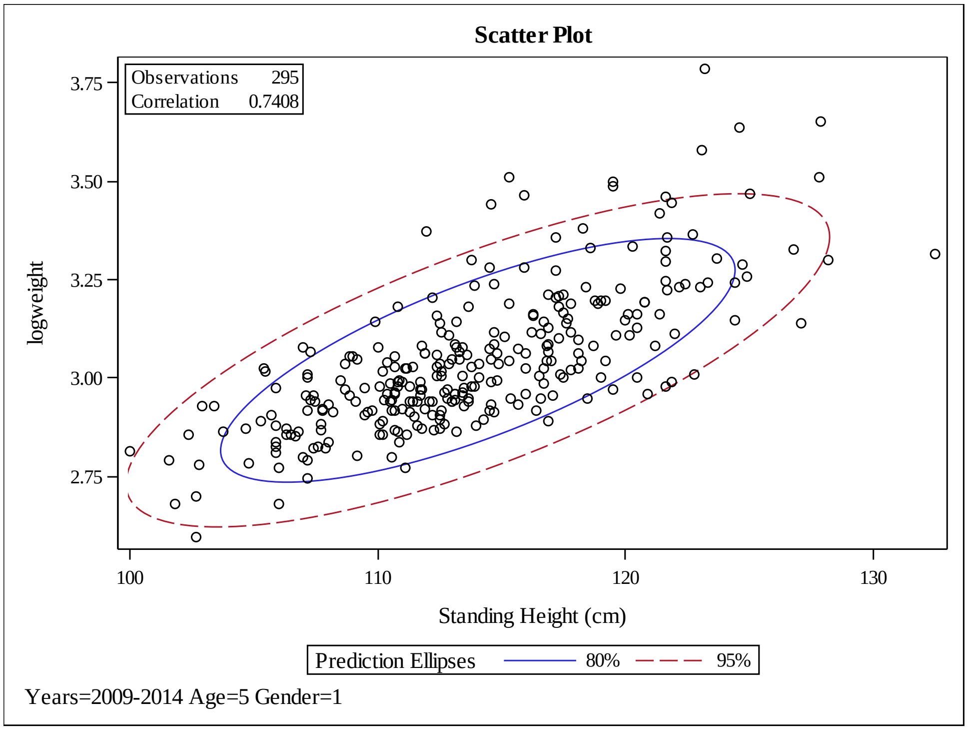

Distribution of BMXHT
Gender = 1
Gender = 2
Age = 20
Age = 25
Percent
0
10
20
30
40
140
150
160
170
180
190
Standing Height (cm)
Curves
Normal
Lognormal(Theta=0 Sigma=EST Zeta=EST)

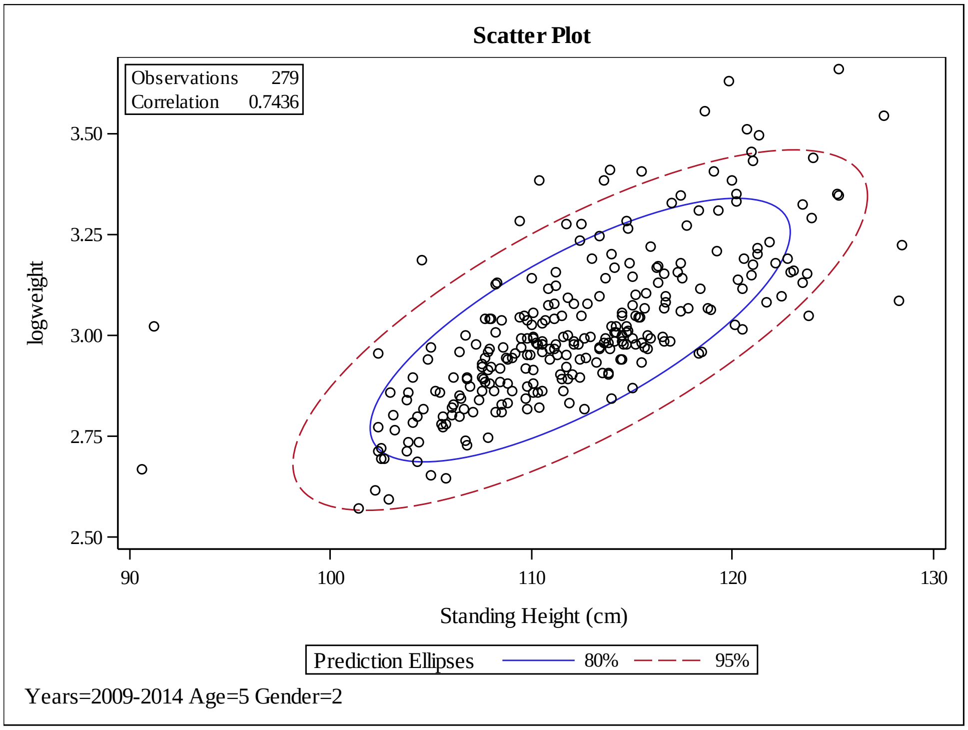

Distribution of BMXHT
Gender = 1
Gender = 2
Age = 30
Age = 40
Percent
0
10
20
30
40
50
60
130
150
170
190
210
230
Standing Height (cm)
Curves
Normal
Lognormal(Theta=0 Sigma=EST Zeta=EST)

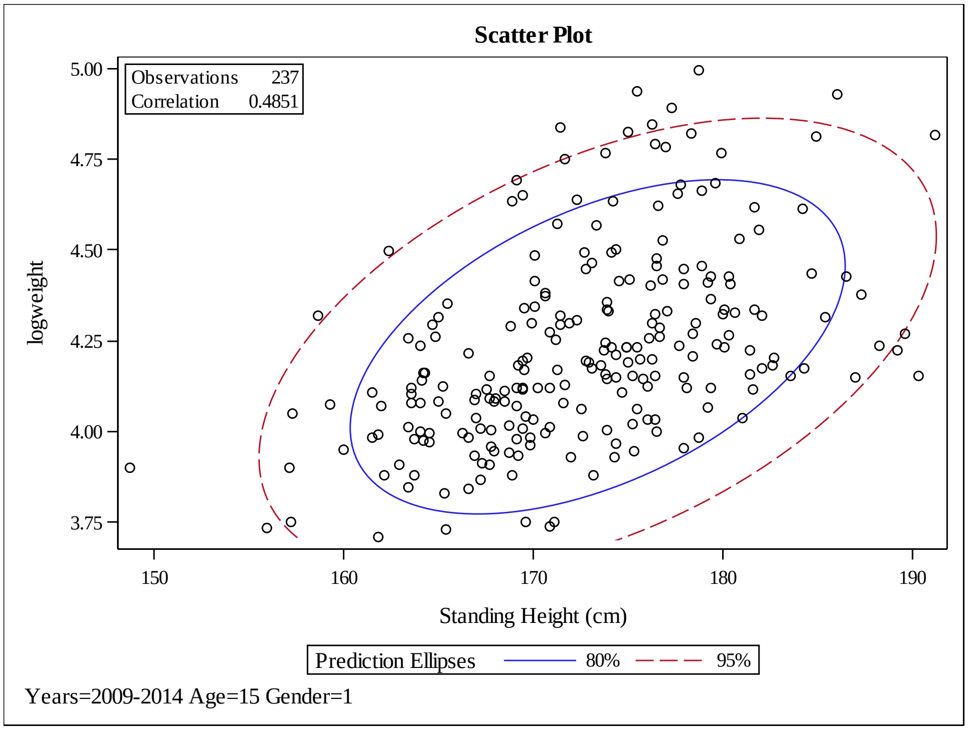

Distribution of BMXHT
Gender = 1
Gender = 2
Age = 50
Age = 60
Percent
0
20
40
60
80
120
140
160
180
200
220
Standing Height (cm)
Curves
Normal
Lognormal(Theta=0 Sigma=EST Zeta=EST)

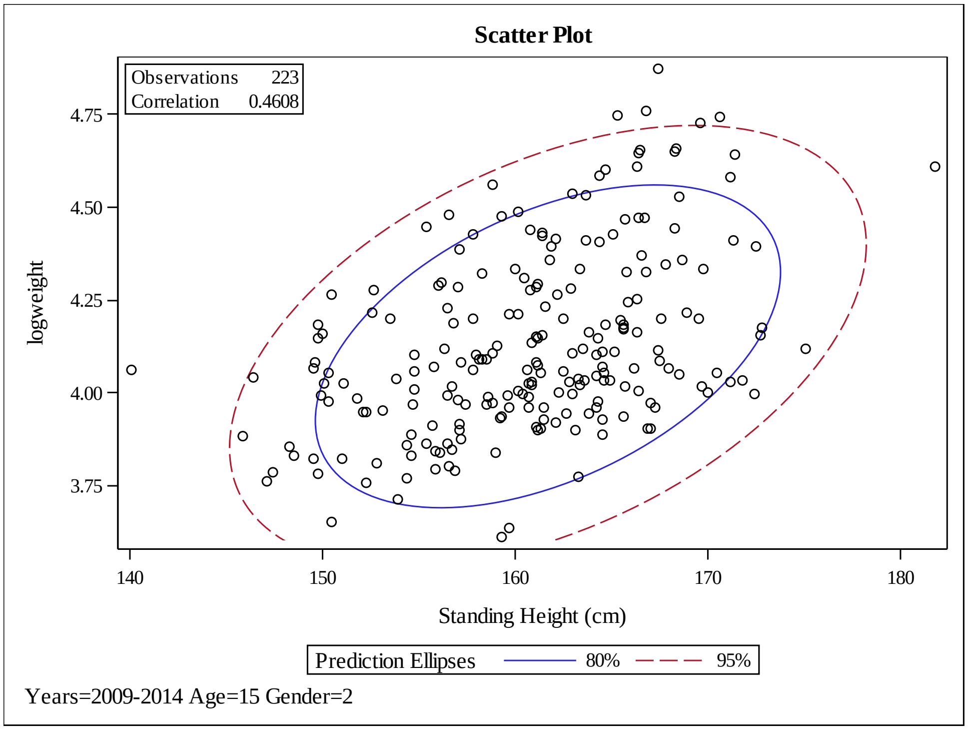

Distribution of BMXHT
Gender = 1
Gender = 2
Age = 70
Age = 79
Percent
0
20
40
60
80
120
140
160
180
200
220
Standing Height (cm)
Curves
Normal
Lognormal(Theta=0 Sigma=EST Zeta=EST)

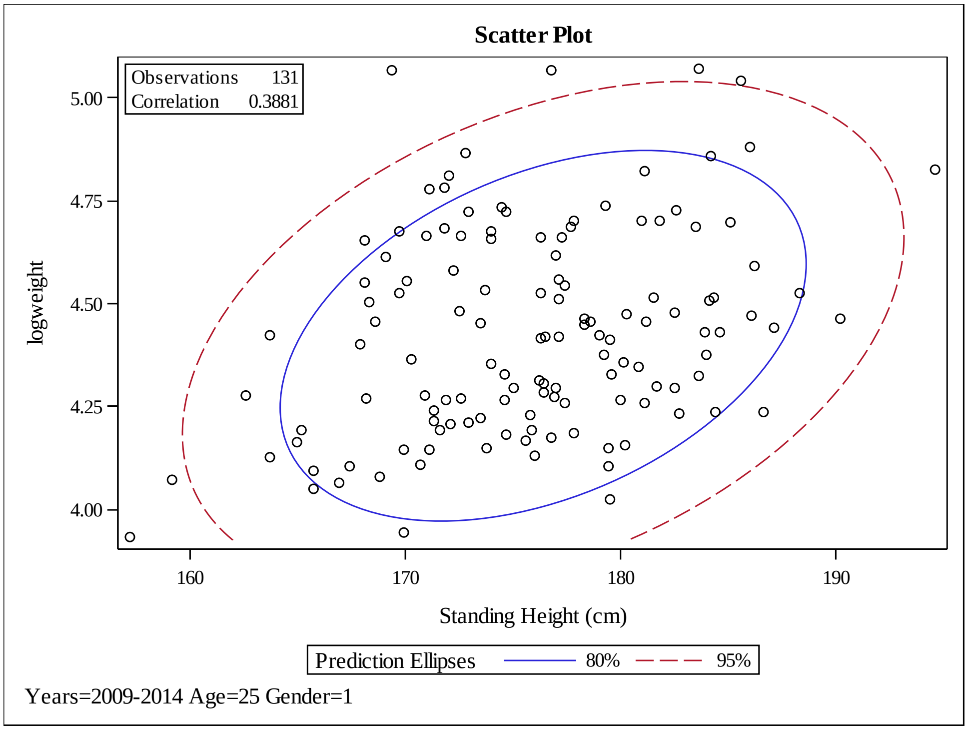
Attachment C. Scatter Plots of Log BW versus HT

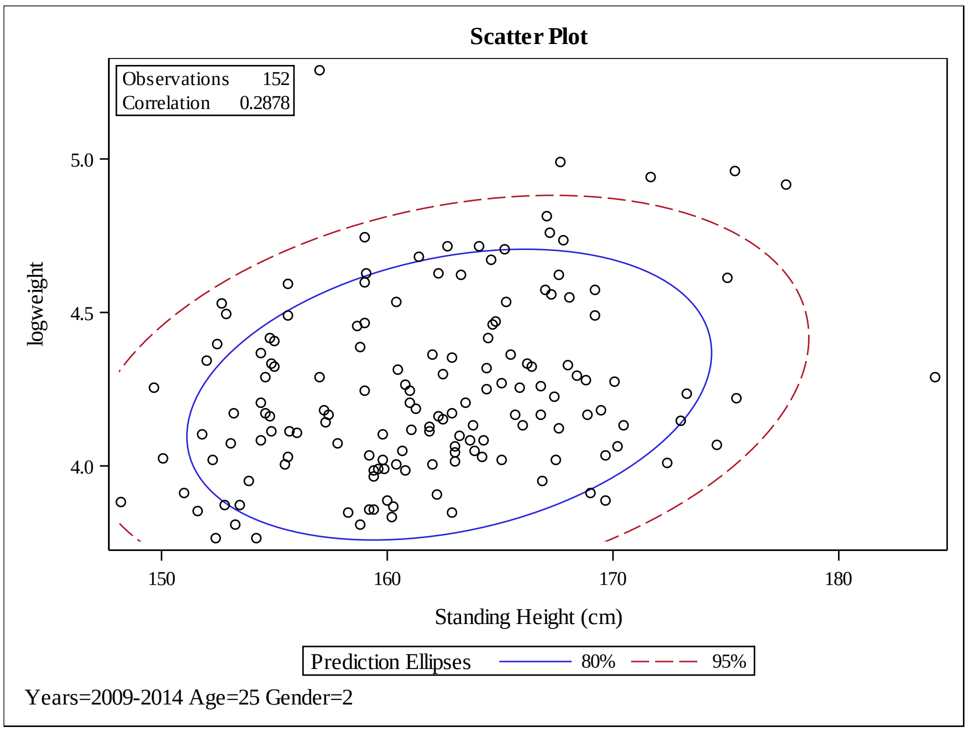


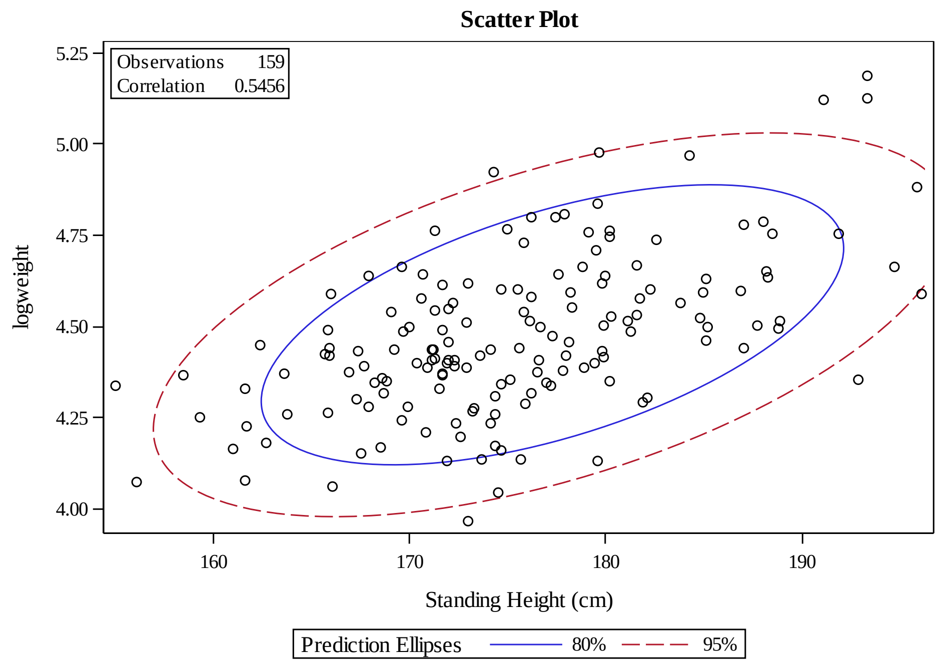



Scatter Plot
Observations 237
Correlation 0.4851
logweight
5.00
4.75
4.50
4.25
4.00
3.75
150
160
170
180
190
Standing Height (cm)
Prediction Ellipses
80%
95%
Years=2009-2014 Age=15 Gender=1

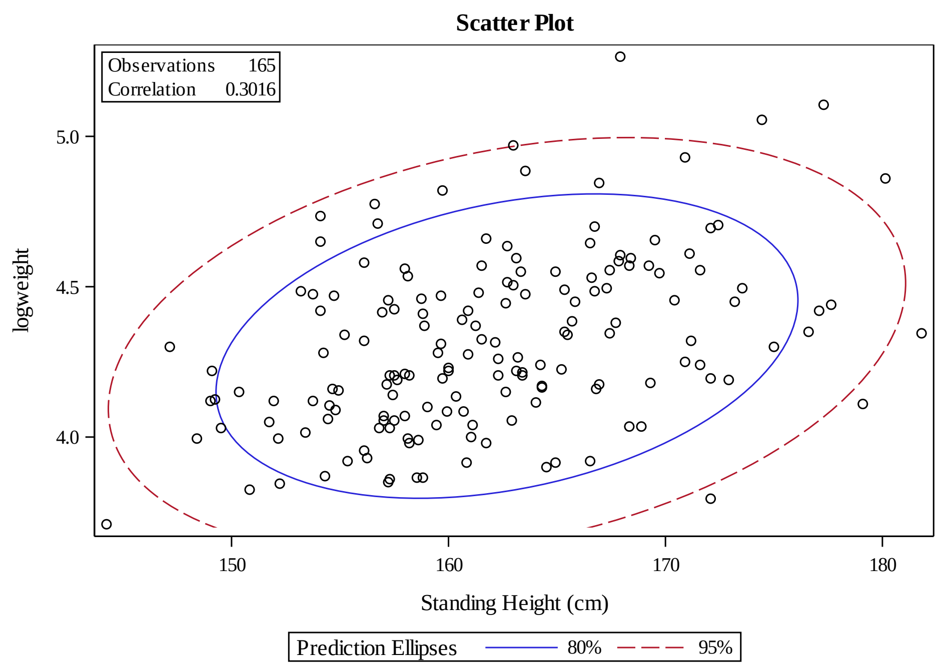

Scatter Plot
Observations 223
Correlation 0.4608
logweight
4.75
4.50
4.25
4.00
3.75
140
150
160
170
180
Standing Height (cm)
Prediction Ellipses
80%
95%
Years=2009-2014 Age=15 Gender=2


Scatter Plot
Observations 131
Correlation 0.3881
5.00
4.75
4.50
4.25
4.00
logweight
160
170
180
190
Standing Height (cm)
Prediction Ellipses 80% 95%
Years=2009-2014 Age=25 Gender=1

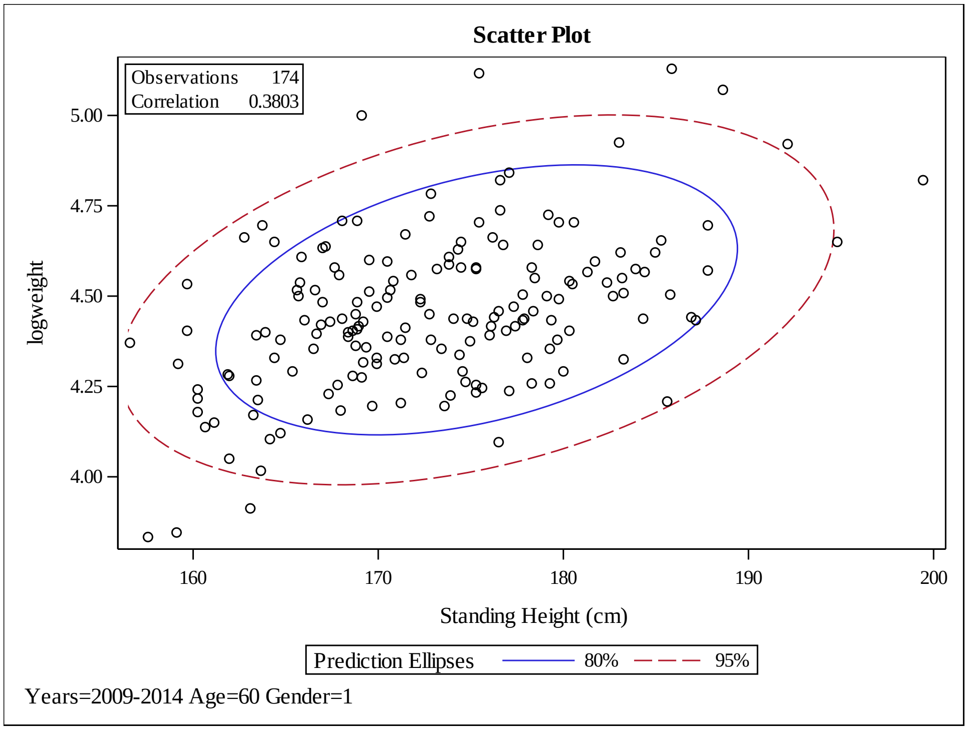

Scatter Plot
Observations 152
Correlation 0.2878
5.0
4.5
4.0
logweight
150
160
170
180
Standing Height (cm)
Prediction Ellipses 80% 95%
Years=2009-2014 Age=25 Gender=2

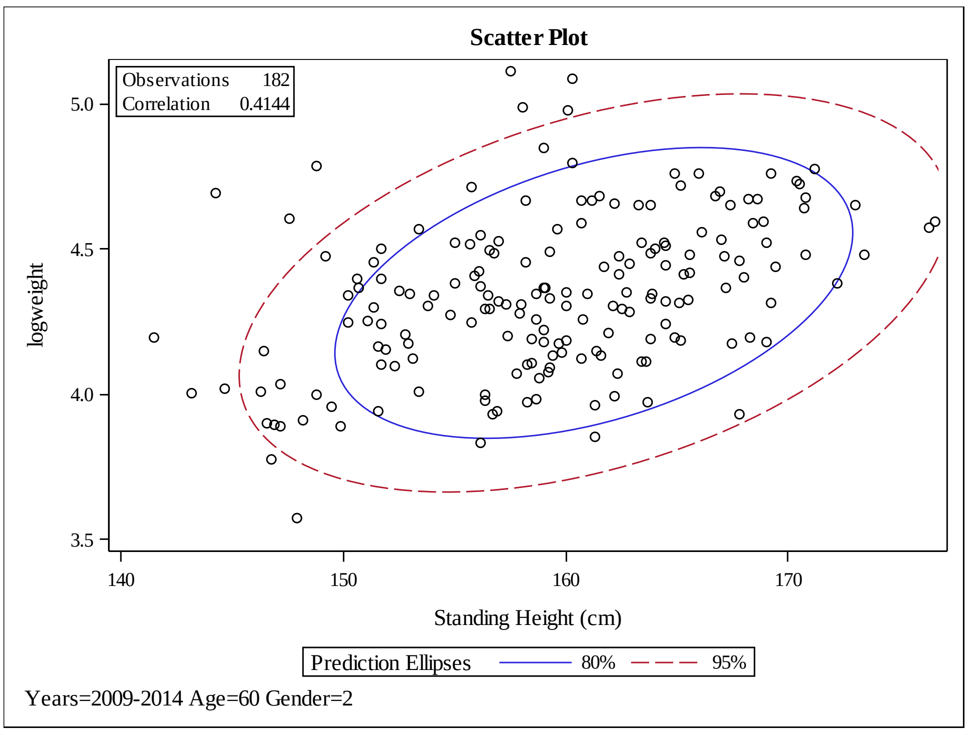


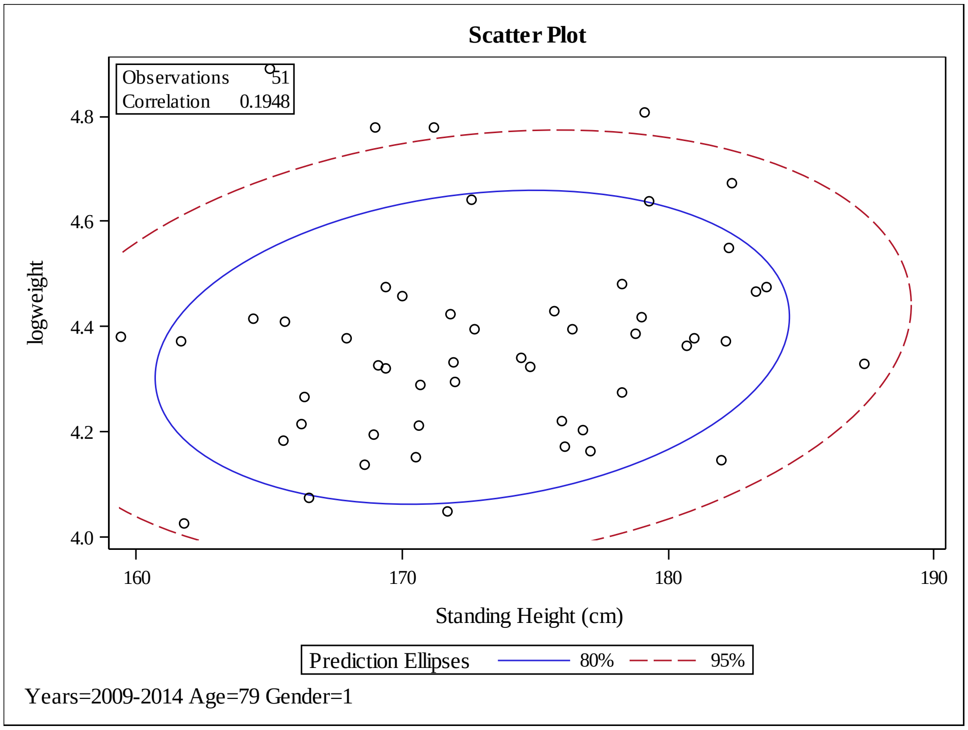
Years=2009-2014 Age=40 Gender=1

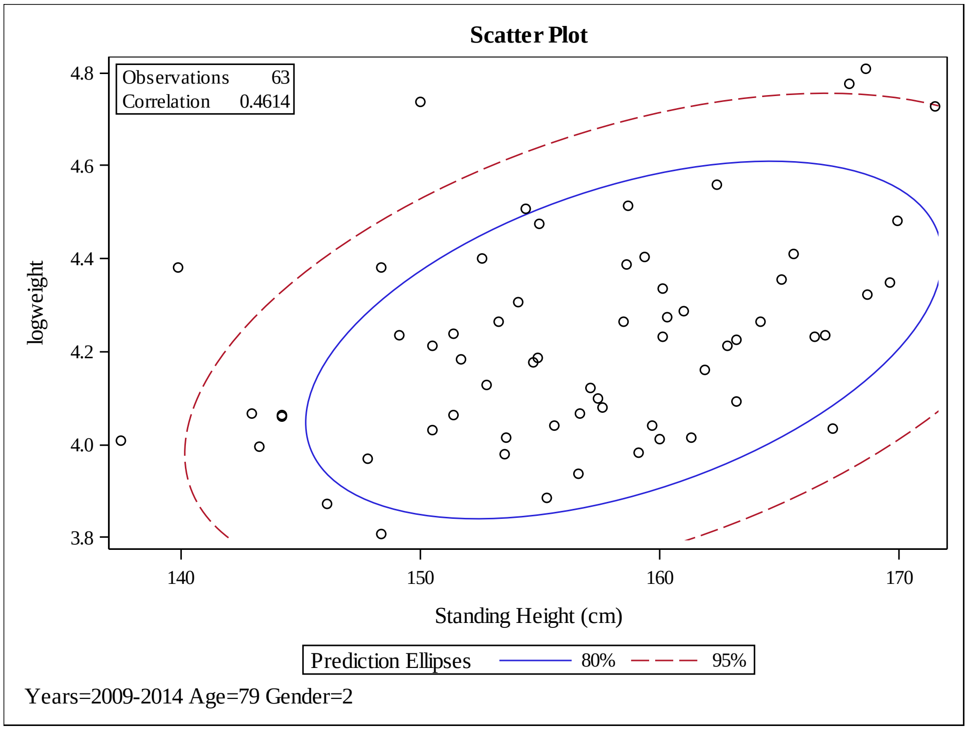


Years=2009-2014 Age=40 Gender=2

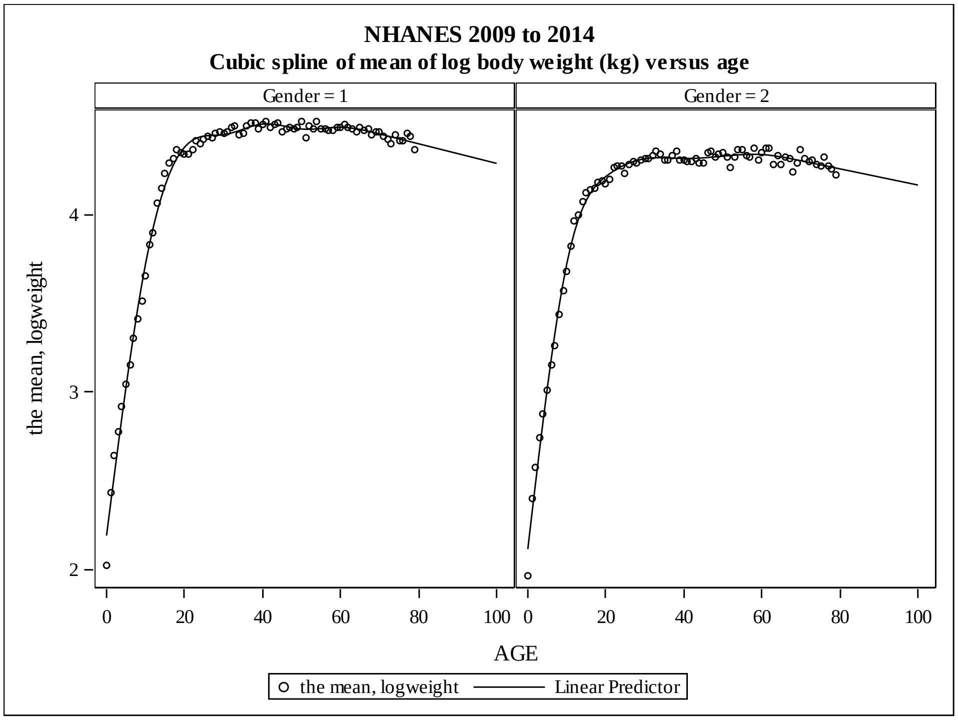

Scatter Plot
Observations 174
Correlation 0.3803
logweight
5.00
4.75
4.50
4.25
4.00
160
170
180
190
200
Standing Height (cm)
Prediction Ellipses 80% 95%
Years=2009-2014 Age=60 Gender=1

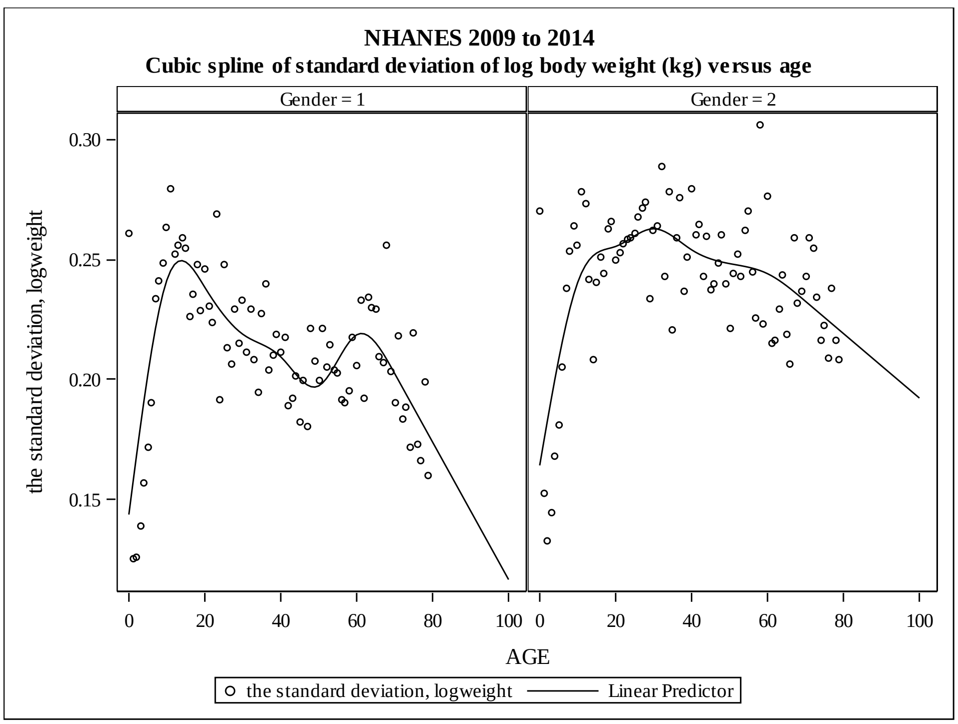

Scatter Plot
Observations 182
Correlation 0.4144
logweight
5.0
4.5
4.0
3.5
140
150
160
170
Standing Height (cm)
Prediction Ellipses 80% 95%
Years=2009-2014 Age=60 Gender=2

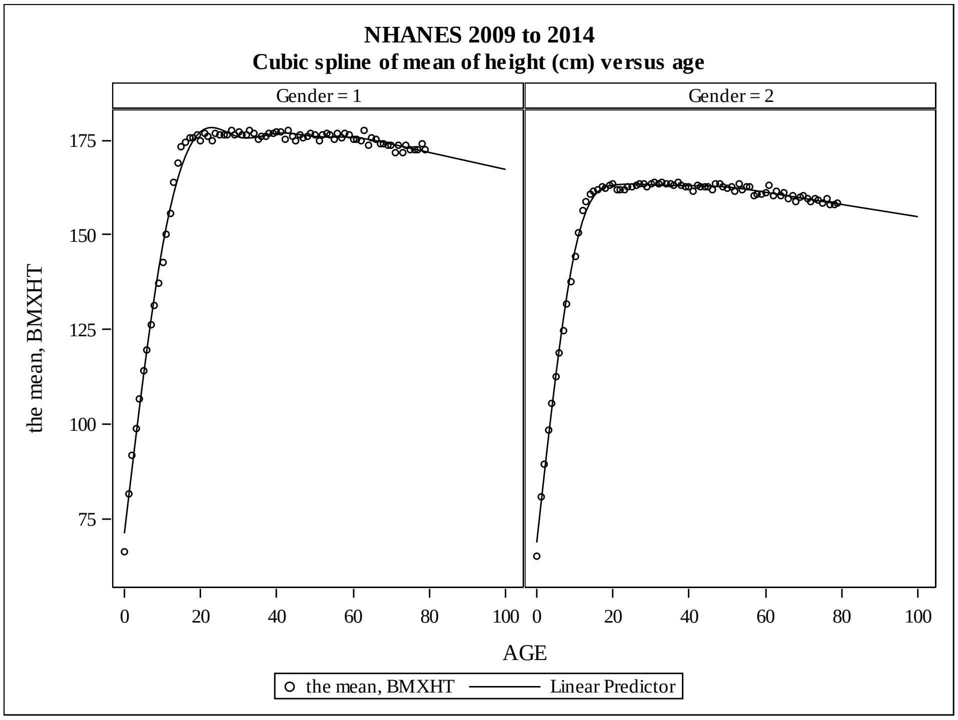

Scatter Plot
Observations 51
Correlation 0.1948
logweight
4.0
4.2
4.4
4.6
4.8
160
170
180
190
Standing Height (cm)
Prediction Ellipses 80% 95%
Years=2009-2014 Age=79 Gender=1

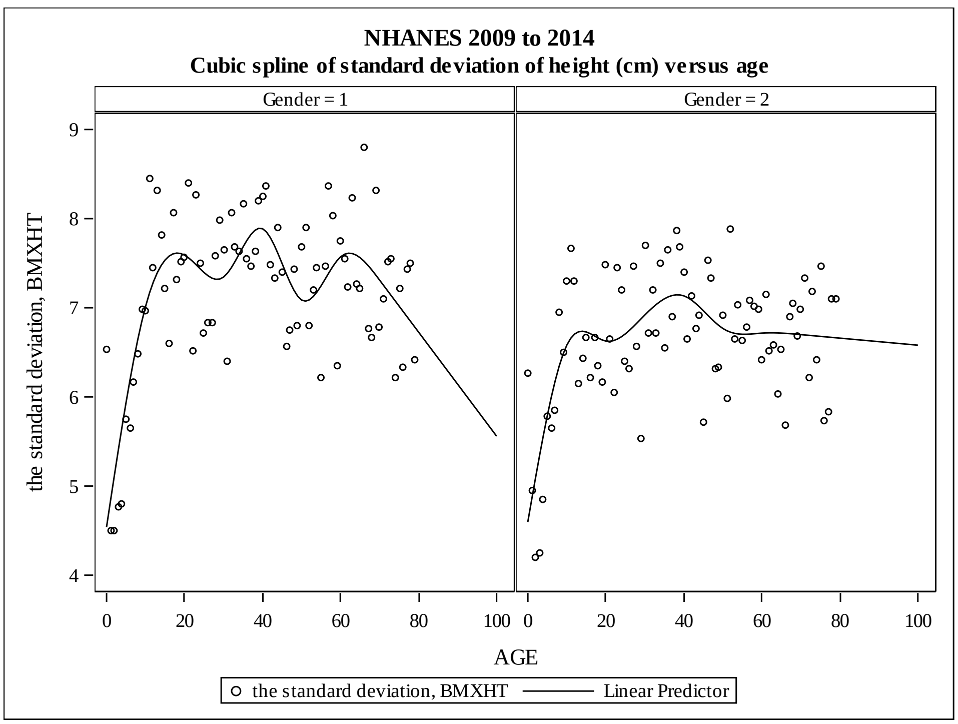

Scatter Plot
Observations 63
Correlation 0.4614
logweight
3.8
4.0
4.2
4.4
4.6
4.8
140
150
160
170
Standing Height (cm)
Prediction Ellipses 80% 95%
Years=2009-2014 Age=79 Gender=2

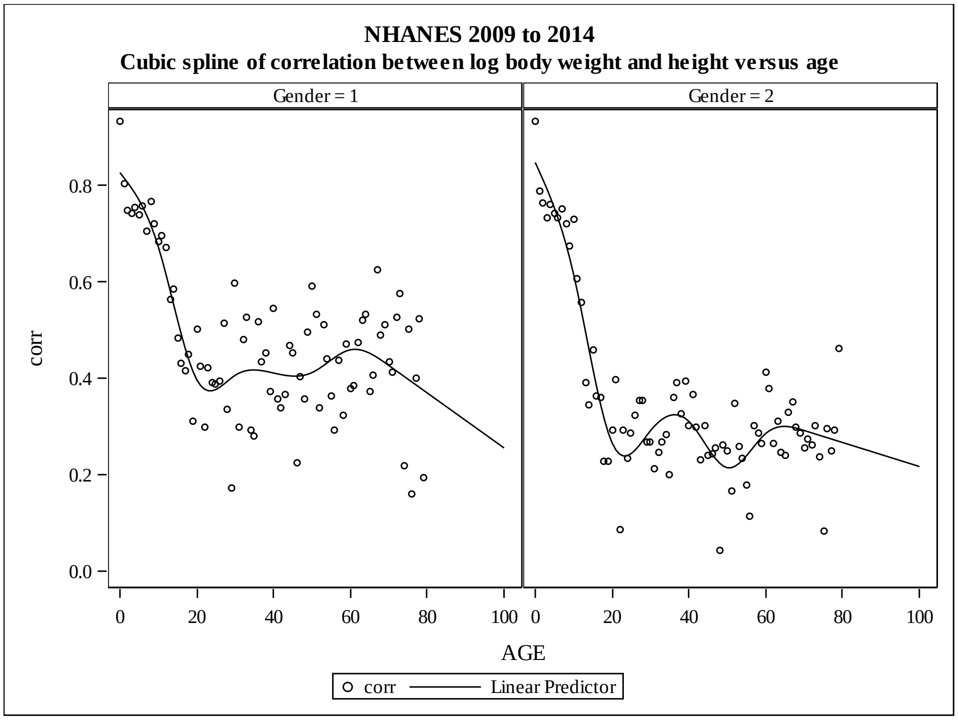
Attachment D. Unsmoothed and Smoothed Values for the Five Joint-distribution Parameters






NHANES 2009 to 2014
Cubic spline of mean of height (cm) versus age
Gender = 1
Gender = 2
the mean, BMXHT
175
150
125
100
75
0
20
40
60
80
100
AGE
the mean, BMXHT
Linear Predictor


NHANES 2009 to 2014
Cubic spline of standard deviation of height (cm) versus age
Gender = 1
Gender = 2
the standard deviation, BMXHT
9
8
7
6
5
4
0
20
40
60
80
100
AGE
the standard deviation, BMXHT
Linear Predictor


NHANES 2009 to 2014
Cubic spline of correlation between log body weight and height versus age
Gender = 1
Gender = 2
corr
0.0
0.2
0.4
0.6
0.8
0
20
40
60
80
100
AGE
corr
Linear Predictor

APPENDIX H

ICF FINAL MEMO: RESTING METABOLIC RATE (RMR)
AND VENTILATION RATE ($\dot{V}_E$) ALGORITHM REFINEMENTS

Memorandum

To:	John Langstaff and Stephen Graham, U.S. EPA OAQPS
From:	Jessica Levasseur, Graham Glen, and Chris Holder, ICF
Date:	February 17, 2017
Re:	WA 4-52 Task 4: RMR and V_E Algorithm Refinements

1. Introduction

Ventilation rate (V_E) and resting metabolic rate (RMR) are two key variables used to assign physiological characteristics to individuals in a simulated population in the U.S. Environmental Protection Agency (EPA) Air Pollutants Exposure model (APEX). These and other simulated aspects of individuals' physiology, combined with population demographics as well as activity data drawn from the EPA Comprehensive Human Activity Database, are used to estimate exposure to air pollutants in APEX (Isaacs, 2008). The current implementation of algorithms used to estimate RMR and V_E in APEX are based on studies that are 30 and 10 years old, respectively (Schofield, 1985; Graham and McCurdy, 2005). The algorithm for V_E also leads to some sharp discontinuities between modeled age groups.

Under this task, ICF ("we") implemented refinements (i.e., technical improvements) to RMR and V_E calculations to improve the usefulness or accuracy of APEX simulations. To complete this task, we conducted multiple literature searches to identify literature relevant to developing appropriate RMR and V_E algorithms. We identified additional sources of data to augment the RMR dataset provided to us by the EPA. We identified no new data on V_E to add to the dataset provided by the EPA.

In this memorandum, we describe these literature searches, the datasets used to develop the updated RMR and V_E algorithms, and the performance in APEX of the updated algorithms for RMR and V_E compared to the existing algorithms. Using updated datasets, we aimed to improve the RMR and V_E algorithms.

Note that all references to "log" or "logarithm" refer to the natural logarithm, not the base-10 logarithm.

2. V_E and RMR Literature Search

In McCurdy (2015), titled "Physiological Parameters and Physical Activity Data for Evaluating Exposure Modeling Performance: a Synthesis," the author expounds upon important factors that influence physiological parameters and affect exposure and dose modeling. He also provided a separate document of "unused references" that contained relevant publications he was unable to fully evaluate in the synthesis.

In focusing on sections containing relevant mentions of V_E and RMR, we identified 321 publications as potentially useful sources of literature that warranted further investigation. We then scrutinized these publication titles and abstracts for particular relevance to RMR or V_E

prediction or to refining the algorithms for RMR or V_E. Of these 321 publications, we identified 53 as potentially relevant for our task.

We identified population gaps within the RMR and V_E datasets initially provided by the EPA, namely women, children, older adults, and obese people (for the V_E dataset) and men and older adults (for the RMR dataset). We focused our literature search on publications specifically relevant to these underrepresented subpopulations.

2.1. RMR

Only 13 publications were relevant to addressing the population gaps present within the RMR dataset. We conducted a cited-references search on these 13 RMR publications, returning all the publications that cite or are cited by these 13 publications. From the RMR cited-references search, we focused on publications that contained each of the following characteristics:

- measured RMR (or an equivalent physiological measurement);
- contained information on body weight, height, and sex; and
- used primary data from at least 200 subjects or defined new predictive equations.

We identified seven publications that had these characteristics. **We acquired new RMR data from one of these publications—the Oxford-Brookes database (Henry, 2005)—adding more than 13,000 unique data points to an RMR dataset provided by EPA.**

2.2. V_E

We conducted a separate literature search for V_E, as requested by the EPA, on those articles published between 2000 and 2010. Conducting a PubMed search on the following search criteria returned 387 publications:

- "Ventilation Rate" OR "VE" AND (Equation/s OR algorithm/s)
- Humans only
- English only.

Assessing these abstracts for new potential sources of data and new potential equations, 16 articles appeared relevant. After acquiring full articles, we identified two as possible sources of data but none had relevant algorithms for V_E prediction. **We were unable to acquire these new datasets for V_E.**

3. Updated RMR Dataset

3.1. Description of Original Dataset

The initial RMR dataset provided to us by the EPA is described in the research report *Analyses of Resting Energy Expenditure (REE) data for US residents* by Kriti Sharma, Thomas McCurdy, and Stephen Graham (no date), which describes a database of 763 individuals ages 4 to 89.

3.2. Description of Oxford-Brookes Database

Published in 2005, Dr. Jaya Henry created the Oxford-Brookes (OB) database that combined data from a variety of sources, resulting in more than 10,000 RMR values. For a detailed summary of the OB database creation, please see Henry (2005) and IOM (2005).

3.3. Merging Datasets

We removed duplicates between the OB database and the initial RMR dataset (provided by the EPA). In addition to information on study author and year of study, this dataset contains information on:

- sex,
- age,
- BM,
- height, and
- RMR.

We deleted observations missing any of the following values: RMR, BM, age, or sex. **The full dataset contains 16,254 observations (9,377 males and 6,877 females).** Of these, 39 males and 33 females were missing reported heights. Therefore, for analyses requiring height (see Section 5), we used a smaller dataset of 16,182 observations (9,338 males and 6,844 females)**.**

4. V_E Dataset

4.1. Description of Dataset

Dr. William Adams of UC Davis constructed the V_E dataset provided to us by the EPA. Graham and McCurdy (2005) also used his data. Dr. Adams collected data from 32 panel studies over 25 years. In addition to information on test exercise parameters, this dataset contains information on:

- sex,
- age,
- BM,
- height,
- oxygen consumption rate (VO_2), and
- V_E.

EPA recommended the removal of four data points for quality-assurance reasons. **The final V_E dataset, with no new data added (none were identified), contains 6,636 observations, with 4,565 males and 2,071 females.**

5. Updated RMR Algorithms

Using the new RMR dataset, and with a goal of improving the RMR algorithm while reducing discontinuities in RMR between age groups, we developed new algorithms for estimating RMR in APEX. The algorithms follow the general format of a multiple linear regression (MLR) model, which is described as:

$$y = \beta_1 x_1 + \beta_2 x_2 + \cdots \beta_n x_n + \alpha + \varepsilon_i(\mu_i, \sigma_i) \quad (1)$$

Where:

- y = variable of interest
- β = coefficient of input variable
- x = input variable
- α = intercept
- ε = residual
- μ = distribution mean
- σ = distribution standard deviation
- n = number of independent regression variables
- i = person-specific index

It is generally known that RMR and BM, as well as RMR and age, are not exactly linearly related; the algorithms developed here use BM, age, and the natural logarithms of BM and (age+1). The "+1" modifier allows APEX to round age upwards instead of downwards to whole years, which is necessary to avoid undefined log(0)[1] values.

To place all the RMR data on an equal footing, we first rounded all ages down to integer values. Instead of dividing the data at preset age boundaries (as was done in the existing APEX algorithm), we repeatedly altered the age boundaries until the residual sum of squares was minimized. Five age groups were sufficient to capture the data for both males and females, though each sex required different age groups. These age groups are shown in Table 1 and Table 2 below, along with the optimal regression parameters (not including height) for each age group and sex. Note that all people over age 99 are treated as 99 years old by APEX and therefore are included in the oldest age groups.

Table 1. Optimal RMR Regression Parameters for Males by Age Group (n = 9,377), Height Not Included

Age Group	n	BM	log(BM)	Age	log(Age)	Intercept	St. Dev.
0–5	625	13.19	270.2	-18.34	131.3	-208.5	69.10
6–13	1355	10.21	260.2	13.04	-205.7	333.4	115.3
14–24	4123	0.207	1078.	115.1	-2794.0	3360.6	161.1
25–54	2531	2.845	729.6	3.181	-191.6	-1067.	178.2
55–99	743	9.291	264.8	-5.288	181.5	-705.9	163.6

Units: RMR = kilocalories/day; BM = kilograms; Age = years

[1] Note that all references to "log" or "logarithm" refers to the natural logarithm, not base-10 logarithm.

Table 2. Optimal RMR Regression Parameters for Females by Age Group (n = 6,877), Height Not Included

Age Group	n	BM	log(BM)	Age	log(Age)	Intercept	St. Dev.
0–5	625	11.94	261.5	-22.31	120.9	-183.6	64.16
6–13	1618	5.296	409.1	40.37	-524.9	392.7	99.43
14–29	2657	0.968	676.9	40.89	-1002.	772.7	143.1
30–53	1346	4.935	355.4	16.28	-896.0	2225.	145.3
54–99	631	2.254	445.9	5.464	-489.9	944.2	124.5

Units: RMR = kilocalories/day; BM = kilograms; Age = years

Input values should be in units of kilograms (kg) for BM and years for age, with the RMR estimate in kilocalories/day (kcal/d). For example, using Equation (1) with information from Table 1, a 20-year-old male weighing 75 kg would be assigned an RMR as follows:

$$RMR = 0.207 \times 75 + 1078 \times \log(75) + 115.1 \times 20 - 2794 \times \log(21) + 3360.6 \\ + 161.1 \times N(0,1)$$

$$RMR = 1826.4 \frac{kcal}{day} + 161.1 \times N(0,1) \text{ (for any 20-year-old male weighing 75 kg)}$$

While the overall r^2 values are fairly high (0.820 males, 0.816 females), the r^2 for particular age groups varies from over 0.9 (for boys and girls ages 0–5 years) to less than 0.6. Transforming RMR, and including height and log(height) as input variables, did not improve overall fit. For adults in particular, a substantial amount of variation remains in the residual error of the new RMR algorithms. To reduce this, more modeling variables would be required than are available in the RMR dataset.

When including height, the optimal regression parameters are as shown in Table 3 and Table 4 for males and females, respectively. The overall r^2 values are 0.815 for males and 0.816 for females when height is included in the regression. These are not appreciably different from the regressions without height. **Therefore, the proposed updates to RMR regressions do not use height.**

Table 3. Optimal RMR Regression Parameters for Males by Age Group (n = 9,338), Height Included

Age Group	n	BM	log(BM)	Age	log(Age)	HT	log(HT)	Intercept	St. Dev.
0–5	596	17.61	106.3	-17.93	87.37	-368.9	676.3	607.6	68.60
6–13	1355	12.64	149.3	30.91	-417.0	-1498.	2151.5	2344.9	115.0
14–24	4123	0.0309	1098.6	114.3	-2777.	31.45	-101.2	3250.7	161.1
25–54	2522	4.692	481.5	2.422	-136.3	1590.	-2014.	-1961.3	176.6
55–99	742	12.60	-108.4	-5.151	170.6	-927.2	2405.	982.6	160.7

Units: RMR = kilocalories/day; BM = kilograms; Age = years; Height = meters

Table 4. Optimal RMR Regression Parameters for Females by Age Group (n = 6,844), Height Included

Age Group	n	BM	log(BM)	Age	log(Age)	HT	log(HT)	Intercept	St. Dev.
0–5	611	21.78	-16.26	-9.014	39.09	-942.8	1259.9	1443.0	61.89

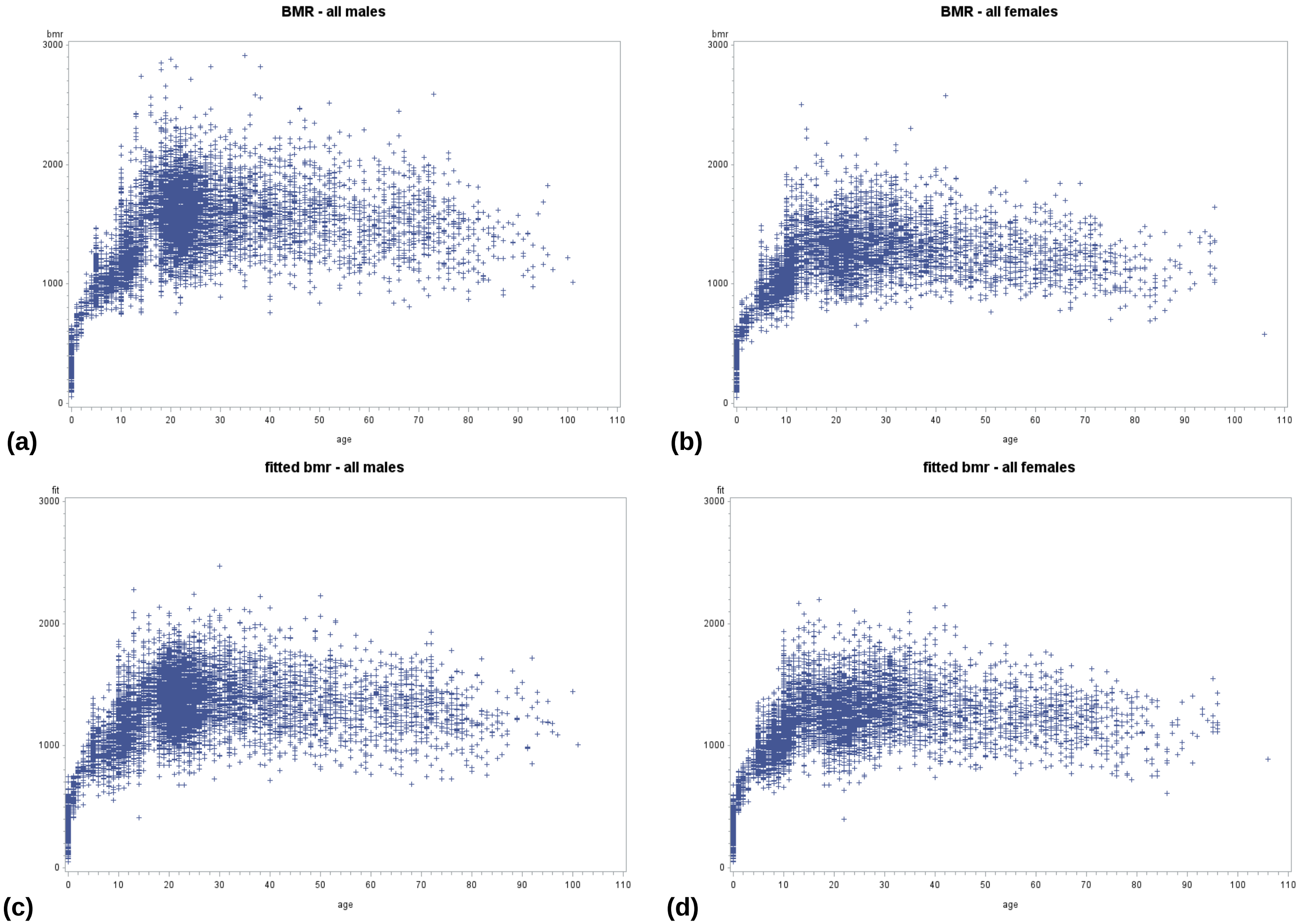
Age Group	n	BM	log(BM)	Age	log(Age)	HT	log(HT)	Intercept	St. Dev.
6–13	1618	7.540	262.8	43.41	-604.3	-338.0	758.7	1209.3	98.85
14–29	2648	4.194	391.6	41.38	-1010.3	152.5	433.1	1298.2	141.1
30–53	1346	6.239	208.5	14.38	-803.3	2854.4	-4066.	-180.9	143.9
54–99	621	3.840	284.9	4.510	-400.1	1782.8	-2274.	-588.6	123.1

Units: RMR = kilocalories/day; BM = kilograms; Age = years; Height = meters

We tried many variations on the above regressions, including changing the age cutpoints, the number of age groups, the list of independent variables, and the transformation of the dependent variable RMR. The SAS program provided in Appendix A contains the code that produces the regressions in Table 1–Table 4 and some of the plots shown below.

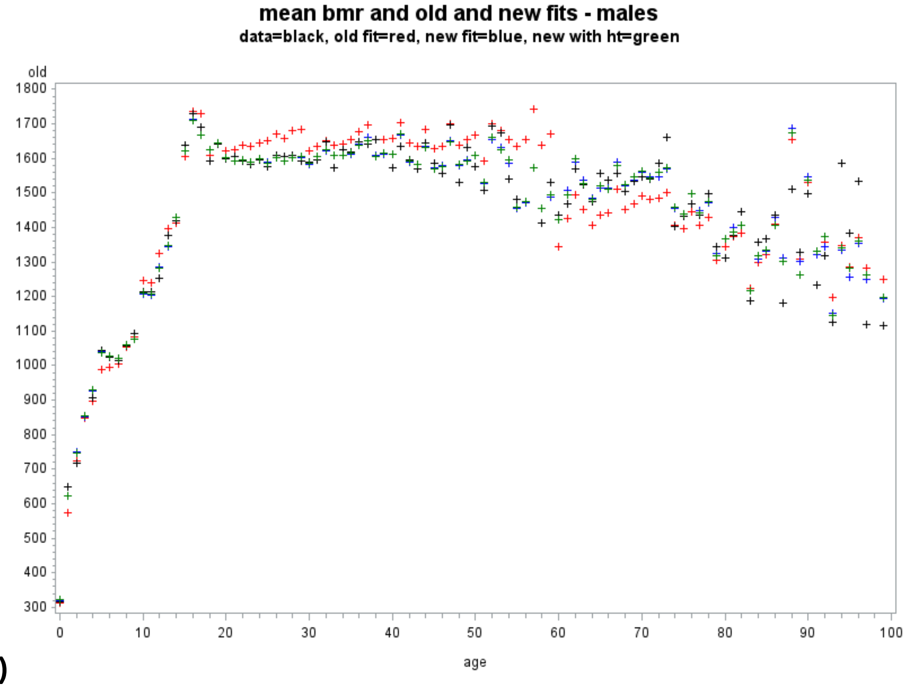
Figure 1 presents scatter plots of observed RMR values (top row) and RMR values predicted by the updated algorithms described above (bottom row), as a function of age. These figures use “BMR” to mean “RMR.” **The updated RMR algorithms have a bias of less than 0.5 percent between observed and predicted values, compared to the existing APEX algorithms which have a bias of 1–2 percent (10–30 kcal/d; smaller bias for females).**

Figure 2 shows the mean RMR values by age: observed (black), predicted by the existing APEX algorithms (red), predicted by the updated algorithms (blue), and predicted by the updated algorithms with height included as an input variable (green; height-related regression parameters not provided in this memorandum). In the red data points (the existing APEX algorithms), a discontinuity is seen between ages 59 and 60, particularly for males. For adults ages 59 and under, the red points are generally higher than the black points (the observed values), whereas the red points are generally below the black points for ages 60 and above. The same effect is seen in females, but the discontinuity is less pronounced. **In the blue data points (the updated algorithms), no sizeable discontinuities are seen at the age group boundaries.** As discussed earlier, the inclusion of height (the green points) does not have a dramatic impact on the fit of the new RMR algorithm.

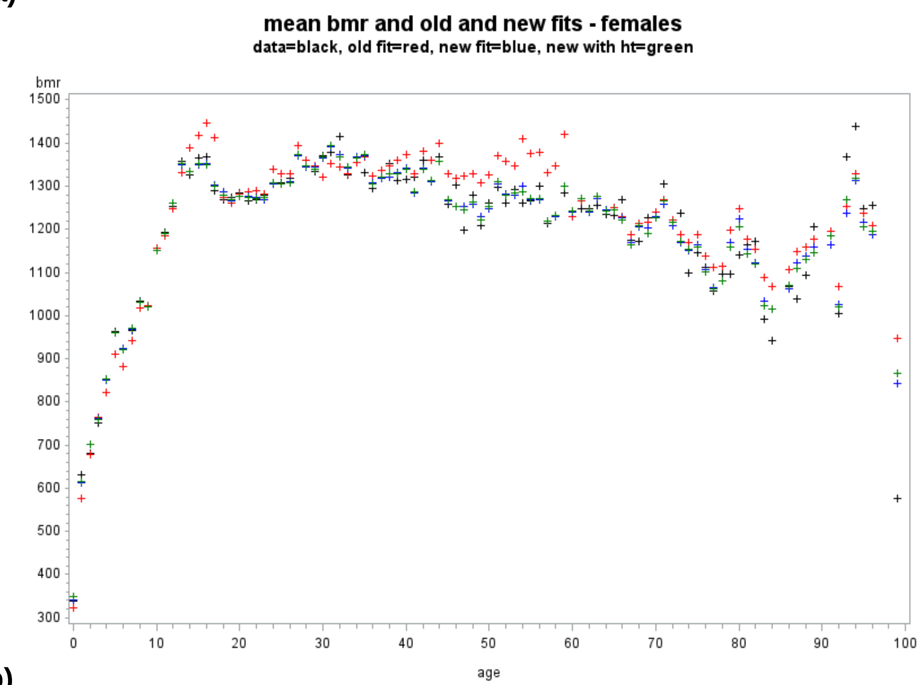


Units: RMR = kilocalories/day, Age = years

Figure 1. Top Row: Observed RMR Values by Age for (a) Males and (b) Females. Bottom Row: Predicted RMR Values by Age for (c) Males and (d) Females using the Updated Algorithms (without height).



(a)



(b)

Units: RMR = kilocalories/day, Age = years

Figure 2. Mean RMR Values by Age: Observed (Black), Predicted by the Existing APEX Algorithms (Red), Predicted by the Updated Algorithms (Blue), and Predicted by the Updated Algorithms with Height Included as an Input Variable (Green), for (a) Males and (b) Females.

6. Updated V_E Algorithm

Using the existing V_E dataset from Graham and McCurdy (2005), we developed updated V_E algorithms for APEX that reduce discontinuities in predicted V_E between age groups and that also utilize maximum VO_2 (VO_2m) as an input. VO_2m is included because ongoing related work on metabolic equivalents of task (MET) values for persons with unusual maximum capacity for work suggests that their MET distributions are modified in a predicable way by their maximum MET (or, equivalently, by VO_2m). One potential limitation of this analysis is that the VO_2m values might not be well characterized for all people in the dataset.

As discussed earlier with Equation 1 above, we aimed to follow the general format of an MLR model. In considering V_E in particular, the available variables for regression are listed in Table 5 below. As discussed later in this section, **we only utilized VO_2 and VO_2m in the updated V_E algorithms**.

Table 5. Summary of Variables Available in the V_E Dataset.

Field	Description
Step	Stage of exercise regimen at a given work level (0.1–13). <1 indicates resting state where 0.1=lay, 0.2=sit, and 0.3=stand (these were not used as they appeared consistently unusual with regard to values observed in the exercising dataset).
Age	Age (y)
BM	Body mass (kg)
Char	Special characteristics of the study subject. 1=trained athlete; 2=trained non-athlete; 3=normally active; 4=sedentary; 5=obese.
ET	Cumulative test time at end of step (min). "."=missing.
Gend	1=females; -1=males
Grd	Percent grade while on treadmill. "."=missing.
HR	Heart rate (b/min) measured during the last minute of each step. "."=missing.
HT	Height (cm)
LBM	Lean body mass (kg)
Mach	Machine used. 1=cycle ergometer; 2=treadmill; "."=missing.
VO_2	Oxygen consumption (L/min, STPD) measured during the last minute of each step
Spd	Treadmill speed (m/min). "."=missing.
STUD	Study number
SUBJ	Study subject identifier
TT	Total time of test (min). "."=missing.
VE	Ventilation (L/min, BTPS) measured during the last minute of each step
VO_2m	Observed VO_2max (L/min, STPD) for the test
Wk	Cycle ergometer setting (W). "."=missing.
ln_ve	log(V_E)
ln_vo2	log(VO_2)
VQ	$V_E \div VO_2$
ln_VQ	log(VQ)
ln_bm	log(BM)
ve_bm	$V_E \div$BM
ln_ve_bm	log(ve_bm)
vo2_bm	$VO_2 \div$BM
ln_vo2_bm	log(vo2_bm)

Note: y = years; kg = kilograms; min = minutes; b/min = beats per minute; cm = centimeters; L = liters; m/min = meters per minute; log = natural logarithm; STPD = standard temperature and pressure, dry; BTPS = body temperature and pressure, saturated.

Out of a total 6,636 observations, 65 had values of VO_2m that were less than values of VO_2. We found that using VO_2m as-is, versus using the maximum between VO_2m and VO_2, made no appreciable difference in estimates of V_E; we therefore used VO_2m as-is.

Each V_E regression took place in two stages. First, all 6,636 data points were used in each regression. Then, all the points that were more than 3 studentized residuals away from the fitted line were removed, and the regression was repeated. This was done to prevent a few outlier points from having undue influence. In this second step, 43 points were rejected though overall they had very little effect on the regression. Note that for a random sample of 6,636 points from a true normal distribution, about 18 would be expected to be more than 3 standard deviations from the mean. The number of outliers was therefore only modestly above what would be expected by chance alone.

The Graham and McCurdy (2005) regressions had four separate age groups (<20, 20–33, 34–60, and 61+) evaluated independently, so discontinuities appear at the age boundaries. Thus, a given person ageing across a boundary would experience a sudden shift in their V_E /VO_2 relationship. Our new analysis uses the same regression equation for all ages, eliminating this issue.

For a given VO_2 level, if VO_2m decreases, then (VO_2/VO_2m) increases, and thus V_E also increases. This relationship eliminated the need to regress upon variables such as age, BM, height, and sex. For example, males on average need less V_E to support a given VO_2, which is captured by their having higher VO_2m. The only variables needed for the new V_E algorithm are VO_2 and VO_2m, both of which are already calculated in APEX.

The actual values of VO_2 and VO_2m are less relevant than the fraction of maximum capacity, represented by $f_1 = VO_2/VO_2m$. f_1 may operate non-linearly (for example, $f_1 = 0.9$ is likely *more* than twice as encumbering as $f_1 = 0.45$). A SAS procedure “Proc Transreg” was used to determine appropriate transformations. This recommended a power of 4 or 5 be used, that is, $y = V_E^{-0.25}$ or $y = V_E^{-0.2}$, when only the variable ln_vo2 was used as the independent variable.

Table 6. Reported r^2 Statistic Based on Transformation of V_E

Transformation of V_E	Variables	tr_r2	ve_r2
2	ln_vo2	0.9479	0.7350
3	ln_vo2	0.9566	0.8779
4	ln_vo2	0.9563	0.8873
5	ln_vo2	0.9544	0.8850
6	ln_vo2	0.9523	0.8821
ln_V_E	ln_vo2	0.9341	0.8561

Note: VO_2 = oxygen consumption rate; ln_vo2 = log(VO_2) = natural log of VO_2; transformation of V_E is V_E^{-N} when N is an integer; ln_V_E= log(V_E); tr_r^2 = r^2 of the transformed response variable,ve_r^2 = r^2 of V_E

Table 6 demonstrates that the reported r^2 for the regression (called tr_r2) of the transformed variable $Y = V_E^{(-1/power)}$ is higher than the r^2 for V_E itself (called ve_r2), but that reflects how well the regression captures the variation in the transformed variable. Because the transformation is intended to “linearize” the data, it is expected that the regression would fit better on the transformed variable. Note that the set of variables that produce the optimal r^2 for the transformed variable sometimes is not the same set that is optimal for ve_r2.

When ln_vo2 is the only independent variable, the best transformation (in terms of ve_r2) is power=4, or $y = V_E^{-0.25}$, as seen in Table 7. Table 7 shows that the addition of age, sex, or height makes little impact on the prediction of V_E. Of these, height is the most effective, but it adds less than 0.01 to r^2. However, the addition of either VO_2m or $f_1 = VO_2/VO_2m$ to the set of independent variables gives a substantial improvement in both tr_r2 and ve_r2. However, note that using f_2 instead of f_1 did not improve the fit.

Table 7. Reported r^2 Statistic for Variables used with $Y=V_E^{-0.25}$

Transformation of V_E	Variables	tr_r2	ve_r2
4	ln_vo2	0.9563	0.8873
4	ln_vo2, age	0.9566	0.8900
4	ln_vo2, sex	0.9578	0.8923
4	ln_vo2, height	0.9596	0.8938
4	ln_vo2, VO_2m	0.9715	0.9213
4	ln_vo2, f_1	0.9721	0.9378
4	ln_vo2, f_2	0.9712	0.9347

Note: VO_2 = oxygen consumption rate; VO_2m = maximum VO_2; ln_vo2 = log(VO_2) = natural log of VO_2; tr_r^2 = r^2 of the transformed variable; ve_r^2 = r^2 of V_E; $f_1 = VO_2/VO_2m$; $f_2 = (VO_2/VO_2m)^2$; transformation of V_E is V_E^{-N} when N is an integer

Once f_1 is added to the list of independent variables, then the optimal transformation of V_E changes. For example, the first line of

Table 8 shows that a power of 5 (that is, $y = V_E^{-0.2}$), now outperforms a power of 4 (see the r^2 values in the second-to-last line of Table 7), whereas the opposite was true in Table 6. The optimal transformation of V_E changes and the optimal set of independent variables depend on each other. Using the ve_r2 statistic as the measure, then for power=5, f_2 provides a better fit that f_1, but that f_3 is worse than f_2. The same is true for power = 6, although all the fits (except for the one using f_1) are better than with power = 5.

Even higher transformation powers can be used, but in practice large powers provide similar results to a log transformation[2]. The last five rows of

[2] The SAS Proc Transreg uses the symbolism power=0 to explicitly indicate a log transformation for the response variable, although since the Tables report values of (-1/power), it would be more correct to call this power = ∞

Table 8 examines using the natural logarithm of V_E as the dependent variable, with the natural logarithm of VO_2 and various powers of (VO_2/VO_2m) as independent variables. Using f_1 or f_2 provides a worse fit with ln_V_E than is obtained with power = 6, but using f_4 provides the best overall fit.

Table 8. Reported r^2 for Combinations of Independent Variables and Transformations of V_E

Transformation of V_E	Variables	tr_r2	ve_r2
5	ln_vo2, f_1	0.9730	0.9402
5	ln_vo2, f_2	0.9729	0.9420
5	ln_vo2, f_3	0.9723	0.9402
6	ln_vo2, f_1	0.9730	0.9397
6	ln_vo2, f_2	0.9734	0.9445
6	ln_vo2, f_3	0.9731	0.9442
6	ln_vo2, f_4	0.9723	0.9427
ln_V_E	ln_vo2, f_1	0.9662	0.9244
ln_V_E	ln_vo2, f_2	0.9714	0.9411
ln_V_E	ln_vo2, f_3	0.9724	0.9466
ln_V_E	ln_vo2, f_4	0.9719	0.9481
ln_V_E	ln_vo2, f_5	0.9711	0.9479

Note: VO_2 = oxygen consumption rate; VO_2m = maximum VO_2; ln_vo2 = log(VO_2) = natural log of VO_2; f_1 = VO_2/VO_2m; $f_N = (VO_2/VO_2m)^N$; transformation of V_E is V_E^{-N} when N is an integer; tr_r^2 = r^2 of the transformed variable; ve_r^2 = r^2 of V_E

Using the log transformation with the independent variables ln_vo2 and $f_4=(VO_2/VO_2m)^4$, Table 9 examines the effects of adding further independent variables; specifically age, gender, and/or height.

Table 9. Various Sets of Independent Variables used to Predict log(V_E)

Transform	Variables	tr_r2	ve_r2
ln_V_E	ln_vo2, f_4	0.9719	0.9481
ln_V_E	ln_vo2, f_4, age	0.9720	0.9477
ln_V_E	ln_vo2, f_4, gender	0.9721	0.9483
ln_V_E	ln_vo2, f_4, height	0.9723	0.9481
ln_V_E	ln_vo2, f_4, age gender height	0.9726	0.9477

Note: VO_2 = oxygen consumption rate; ln_vo2 = log(VO_2) = natural log of VO_2; tr_r^2 = r^2 of the transformed variable; ve_r^2 = r^2 of V_E; $f_4 = (VO_2/VO_2m)^4$

In all cases, the ve_r2 is unchanged to three decimal places, being 0.948 in all cases. Hence, the recommendation is to use the simplest version of these regressions, as seen in Equation (2) below.

$$VE2 = e^{(3.298 + 0.7935 \times ln\_vo2 + 0.53845 \times (VO_2 \div VO_2 m)^4 + 0.1253 \times N(0,1))} \qquad (2)$$

The following two figures show all 6,636 data points from the V_E dataset. Figure 3 shows measured V_E and measured VO_2. Figure 4 shows predicted V_E ("VE2") and measured VO_2, where VE2 is given by Equation (2) (with an r^2 of 0.948, as shown in Table 9) which is based on the V_E dataset with outliers removed (this is *not* the final updated V_E algorithm, as noted later in this section).

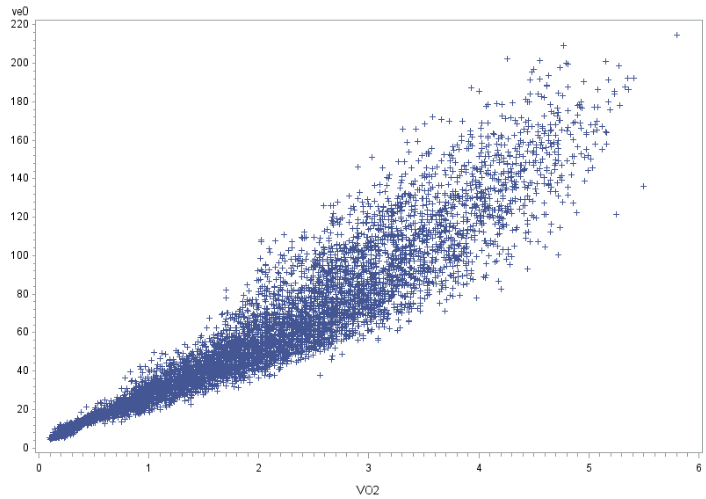


Figure 3. Measured VO_2 and Measured V_E, from the V_E dataset

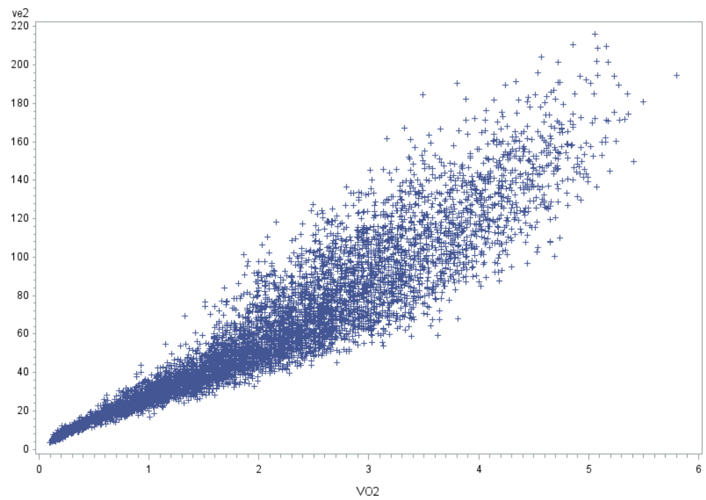


Figure 4. Measured VO_2 and Predicted V_E Using the Updated Algorithm (i.e., VE2 shown in Equation 2).

As can be seen in the figures, **predicted and observed values of V_E are very close.**

In concordance with a request from the EPA WAM, we developed a mixed-effects regression (MER) in addition to the above MLR. MER separates residuals into within-person (ew) and between-person (eb) effects, known as intrapersonal and interpersonal effects, respectively. This analysis, using the same independent variables and the same V_E dataset discussed above yields another V_E algorithm. **This algorithm, shown below, is the final version of the updated V_E algorithm to be incorporated into APEX.**

$$Pred\_VE = e^{(3.300 + 0.8128 \times ln\_vo2 + 0.5126 \times (VO_2 \div VO_2m)^4 + N(0,eb) + N(0,ew))} \quad (3)$$

N(0,eb) is a normal distribution with mean zero and standard deviation eb=0.09866 meant to capture *inter*personal variability, which is sampled once per person. *N(0,ew)* is an *intra*personal residual with standard deviation of ew=0.07852, which is resampled daily due to natural *intra*personal fluctuations in V_E that occur daily.

Differences between Equations (2) and (3) may be due to the fact that some of the persons in the dataset had different numbers of observations. The mean, median, and mode were all seven observations per person, with a range from one to 13. With regard to implementation in APEX, the cause of the interpersonal variability may not be necessary to determine. It is sufficient to specify the size of the two error terms, one sampled once per person and the other sampled once per day.

Ultimately, the EPA WAM chose Equation (3) to implement in APEX due to its increased ability to account for inter- and intra-personal effects. **The resulting r^2 for V_E (0.94) is a substantial improvement over the existing V_E regressions in APEX (where r^2 was 0.892–0.925), with a large reduction in discontinuities of V_E between ages.**

7. Effect of Updated Algorithm(s) on Simulated Exposure

The updated RMR algorithm is based on an MLR with coefficients shown in Table 1 and Table 2. The updated V_E algorithm is shown in Equation (3).

The existing RMR algorithm in APEX (in units of kilocalories/minute [kcal/min]) is:

$$RMR = 0.166 \times \left[RMR_{slope} \times BM + RMR_{int} + RMR_{err}\right] \quad (4)$$

Where:

0.166	=	the conversion factor for converting megajoules (MJ)/d to kcal/min
RMR_{slope}	=	slope of the regression equation (MJ/(d-kg))
RMR_{int}	=	intercept of the regression equation (MJ/d)
RMR_{err}	=	variation in the regression equation (MJ/d)

The existing V_E algorithm in APEX (in units of milliliters/minute [mL/min]) is:

$$V_E = \left(1{,}000\,\frac{mL}{L}\right) \times BM \times \exp\left(V_{Einter} + V_{Eslope} \times ln(VO_2) + Z \times V_{Eresid}\right) \quad (5)$$

Where:

V_{Einter}	=	intercept of the regression equation
V_{Eslope}	=	slope of the regression equation
Z	=	random number from normal distribution

V_{Eresid} = variation in the regression equation

And where VO_2 (in units L/min/kg) is:

$$VO_2 = \frac{MET \times ECF \times RMR}{BM} \tag{6}$$

Where:

ECF = energy conversion factor (L O_2/kcal)

We compared the effects of the existing and updated RMR and V_E algorithms using a sample of 1000 persons, ages 0 to 95, run for one year each (taken from an APEX run for ozone, in 2010, in the Los Angeles area). Four runs were made: R1V1 is the combination of old RMR and old V_E algorithms; R2V1 uses the new RMR and old V_E algorithms; R1V2 uses the old RMR and new V_E algorithms; and finally, R2V2 uses both new algorithms. Each run produced a sample of 1000 RMR values (one per person), and 8,760,000 V_E values (one per hour, per person).

The RMR results did not vary when just the V_E method was changed. This was expected, because APEX calculates RMR first. The V_E calculation is affected by any change in RMR. Statistics comparing the old and new RMR algorithms are presented in Table 10. The new RMR algorithm produces slightly lower values across the board, with larger decreases at the higher end of the range. Even then, these differences are below 4 percent. There are fewer extreme values using the new algorithm, resulting in a smaller standard deviation.

Table 10. RMR Value Statistics (kcal/min) for 1000 Persons, Using Old and New RMR Algorithms

Statistic	Old RMR	New RMR	% Change
Mean	1.065	1.040	- 2.4 %
Standard deviation	0.292	0.275	- 5.8 %
10^{th} percentile	0.709	0.702	- 1.0 %
Median	1.057	1.034	- 2.2 %
90^{th} percentile	1.443	1.390	- 3.7 %

The V_E data below have been analyzed in two ways. First, statistics on the full set of 8,760,000 V_E values are generated. When comparing the same V_E algorithm and varying RMR algorithms, the old V_E algorithm had a drop of 2 percent in mean V_E when switching to the new RMR, and the new V_E algorithm had a similar drop of 1.5 percent (not shown in a table here). These are somewhat smaller than the drop in mean RMR of 2.4 percent.

Focusing on the new RMR algorithm, a comparison of V_E statistics from the R2V1 and R2V2 runs is shown in Table 11, using all 8,760,000 V_E values. The high-end V_E values changed very little between the old and new V_E algorithms (by 0.5 percent), but the new algorithm predicts higher values at lower V_E levels (by 17.6 percent), resulting in an increase by 6 percent in mean values. These values are effectively time-weighted, so sleeping V_E accounts for about one-third of the set (that is, at rest or below). By contrast, the Adams dataset was concerned almost solely with activities above resting levels. Hence, the regression based on the Adams dataset is being extrapolated to sleeping as an activity. One would therefore expect that the new V_E

algorithm would be more robust for the higher activity levels. Note that the new V_E algorithm has a smaller standard deviation than the old method (by 11.6 percent), resulting in fewer extreme values.

Table 11. V_E Value Statistics (mL/hr) for 8,760,000 Person-hours, Using the New RMR Algorithm with the Old and New V_E Algorithms

Statistic	Old V_E	New V_E	% Change
Mean	19581	20763	+ 6.0 %
Standard deviation	10375	9172	- 11.6 %
10^{th} percentile	8778	10319	+ 17.6 %
Median	17422	19391	+ 11.3 %
90^{th} percentile	33042	32887	- 0.5 %

The second type of analysis is to examine the change in mean V_E per person, and the change in the 90^{th} percentile of each person's V_E values. First, the 1000 personal means (over the year) and 1000 personal 90^{th} percentiles are calculated. Table 12 shows modest increases (in the range of 6 percent) in person-mean V_E values when using the new V_E algorithm, with a 1.8-percent increase in standard deviation. Table 13 shows that the 90^{th} percentile for each person (that is, the V_E level that one exceeds for 2.4 hours per day, on average) has changed relatively little between the old and new algorithms. The mean has dropped 2 percent, but the standard deviation dropped by 9.1 percent because the upper tail does not extend as far as before.

Table 12. Population Statistics on Personal Mean V_E (mL/hr), Using the New RMR Algorithm with the Old and New V_E Algorithms

Statistic	Old V_E	New V_E	% Change
Mean	19581	20763	+ 6.0%
Standard deviation	6187	6296	+ 1.8%
10^{th} percentile	12236	12843	+ 5.0%
Median	18955	20504	+ 8.2%
90^{th} percentile	27822	29164	+ 4.8%

Table 13. Population Statistics on Personal 90^{th} Percentile of V_E (mL/hr), Using the New RMR Algorithm with the Old and New V_E Algorithms

Statistic	Old V_E	New V_E	% Change
Mean	28017	27445	-2.0%
Standard deviation	11094	10087	-9.1%
10^{th} percentile	14205	14415	1.5%
Median	27026	27339	1.2%
90^{th} percentile	42572	40775	-4.2%

In summary, in comparing the updated APEX algorithms for RMR and V_E to the existing algorithms:

- Average RMR decreases with the updated RMR algorithms, though remains within 3 percent of RMR predicted by the existing algorithm.
- As expected, the updated V_E algorithm has no effect on predicted RMR.

- The updated RMR algorithm impacts V_E predictions less when utilizing the updated V_E algorithm; this impact is greater at the lower end of estimated V_E values.
- The upper end (90^{th} percentile) of predicted V_E values are similar between the existing and updated V_E algorithms. This appears to be due to two partially cancelling effects: the population 90^{th} percentile of the personal means increased 4.8 percent, but the population 90^{th} percentile of the personal 90^{th} percentiles decreased 4.2 percent.
- The lower end of predicted V_E values is moderately higher with the updated V_E algorithm than with the existing V_E algorithm (a 17.6-percent change in the 10^{th} percentile, which corresponds to sleeping V_E)
- Both the updated and existing V_E algorithms predict V_E values exceeding 100,000 mL/min for roughly 1 in every 65,000 person-hours, which was the hard-coded maximum for V_E in APEX. Note that a switch has been added to the APEX Control Options File to enable or disable the maximum upper limit. This was disabled for the current comparison runs, because truncation of the two tails at the same point would cause the two distributions to look more similar than they otherwise would.

8. Summary Discussion and Next Steps

Through extensive literature searches for both RMR and V_E algorithms, as well as through augmentation of the RMR dataset, ICF has improved upon the RMR and V_E physiological algorithms within the APEX model. These updated algorithms perform better than the existing algorithms in APEX, with reduced discontinuities between APEX age groups and better fits to the measured datasets. ICF has created “switches” within the APEX Control Options File that allows users to choose between the available RMR or V_E algorithms. The coding required to completely replace the older algorithms can be done quickly at EPA’s request.

9. References

Graham, S. and McCurdy, T. (2005) A NERL Internal Research Report: Revised ventilation rate (VE) equations for use in inhalation-oriented exposure models. EPA/600/X-05/008.

IOM. (2005). Dietary Reference Intakes for Energy, Carbohydrate, Fiber, Fat, Fatty Acids, Cholesterol, Protein and Amino Acids. Panel on Macronutrients, Panel on the Definition of Dietary Fiber, Subcommittee on Upper Reference Levels of Nutrients, Subcommittee on Interpretation and Uses of Dietary Reference Intakes, and the Standing Committee on the Scientific Evaluation of Dietary Reference Intakes, Food and Nutrition Board. U.S. Institute of Medicine. National Academies Press. URL: http://www.nationalacademies.org/hmd/Activities/Nutrition/SummaryDRIs/DRI-Tables.aspx. Table: Doubly Labeled Water Data Set: 10_DLW_Database.xls (downloaded 8/24/16)

Henry, CJK. (2005). Basal metabolic rate studies in humans: measurement and development of new equations. Public Health Nutrition. 8(7A): 1133-1152.

Isaacs, K. (2008). Estimating ventilation in human exposure models: Summary. Internal Memorandum. July 28, 2008.

McCurdy, T. Physiological Parameters and Physical Activity Data for Evaluating Exposure Modeling Performance: a Synthesis. U.S. Environmental Protection Agency, Washington, DC, EPA/600/R-15/175, 2015.

Schofield, WN (1985). Predicting basal metabolic rate, new standards and review of previous work. Human Nutrition: Clinical Nutrition. 39C(Supp. 1): 5-41.

```
* Written by WGG at ICF, last revised on October 21, 2016

data raw;
 infile "C:/main/APEX/WA452/exercise/from_Jess/newrmr_JL_30aug16.csv"
firstobs=2 dsd dlm=',';
 length sex $1 author $20 type $6 citation $80 study $40;
 input sex author type age bmr ht bm citation year study recno;
 yage = floor(age);
run;


data good bad all;
 set raw;
 logbm = log(bm);
 logbmr = log(bmr);
 if sex="M" then gender= 1;
 if sex="F" then gender=-1;
 bad = 0;
 if age=. then bad=1;
 if bmr=. then bad=1;
 if bm=. then bad=1;
 if gender=. then bad=1;
 if ht=. then bad=1;
 age0 = age;
 age = floor(age);
 if age>99 then age=99;
 logage = log(1+age);
 logage0 = log(1+age0);
 invage = 1/(1+age);
 bmage = bm*age;
 bmcage = bm*(1+age);
 bmlage = bm*logage;
 loght = log(ht);
 if bad=0 then output good; else output bad;
 output all;
run;

data males females;
 set all;
 if gender= 1 then output males;
 if gender=-1 then output females;
run;




axis1 order = 0 to 3200 by 200;
title 'RMR: All males';
proc gplot data=males;
 plot bmr*age /VAXIS=axis1;
run; quit;

title 'RMR: All females';
```

```
proc gplot data=females;
 plot bmr*age /VAXIS=axis1;
run; quit;




axis1 order = 0 to 3200 by 200;
axis2 order = 0 to 180 by 10;
title 'RMR vs BM: All males';
proc gplot data=males;
 plot bmr*bm /VAXIS=axis1 HAXIS=axis2;
run; quit;
title 'RMR vs BM: All females';
proc gplot data=females;
 plot bmr*bm /VAXIS=axis1 HAXIS=axis2;
run; quit;

axis1 order = 0 to 3200 by 200;
axis3 order = 0 to 2.0 by 0.1;
title 'RMR vs BM: All males';
proc gplot data=males;
 plot bmr*ht /VAXIS=axis1 HAXIS=axis3;
run; quit;
title 'RMR vs BM: All females';
proc gplot data=females;
 plot bmr*ht /VAXIS=axis1 HAXIS=axis3;
run; quit;

proc sort data=males; by yage; run;
proc means data=males noprint;
 by yage;
 var bmr;
 output out=m1 n=n mean=mean std=std min=min max=max;
run;
proc print data=m1; run;





%macro c(gen,num,test,vars);
%let last=0;
%do i=1 %to &num;
 %let j=%scan(&test,&i);
 %put i=&i j=&j;
 title "&gen._&last.-&j";
 data a;
 %if &gen=M %then set males;;
 %if &gen=F %then set females;;
       lo = symgetn("last");
       hi = &j;
 if (age>=lo and age<hi);
       * if ht=. then delete;
 run;
```

```
 proc reg data=a;
 model bmr=&vars /vif;
 output out=z
 p =predicted
 residual=residual
 rstudent=rstudent;
 run; quit;
 data _null_;
 set z end=eof;
       retain pertot 0;
       retain errtot 0;
 pertot = pertot + 1;
 errtot = errtot + residual**2;
       if (eof) then do;
 call symput("pertot&i",trim(left(pertot)));
            call symput("errtot&i",trim(left(errtot)));
 end;
 run;
 %let last = &j;
%end;
%let pertot = 0;
%let errtot = 0;
%do i=1 %to &num;
 %let pertot = %sysevalf(&pertot+&&pertot&i);
 %let errtot = %sysevalf(&errtot+&&errtot&i);
%end;
%put test = &test;
%put pertot = &pertot;
%put errtot = &errtot;
%mend;

%c(F,5,6 14 30 54 100,bm logbm age logage); * err = 11052 best;
%c(M,5,6 14 25 55 100,bm logbm age logage); * err = 22753 best;


%macro d(gen,num,test,vars);
%let last=0;
%do i=1 %to &num;
 %let j=%scan(&test,&i);
 %put i=&i j=&j;
 title "&gen._&last.-&j";
 data a;
 %if &gen=M %then set males;;
 %if &gen=F %then set females;;
       lo = symgetn("last");
       hi = &j;
 if (age>=lo and age<hi);
       if ht=. then delete;
 run;
 proc reg data=a;
 model bmr=&vars /vif;
 output out=z
 p =predicted
 residual=residual
 rstudent=rstudent;
 run; quit;
```

```
 data _null_;
 set z end=eof;
       retain pertot 0;
       retain errtot 0;
 pertot = pertot + 1;
 errtot = errtot + residual**2;
       if (eof) then do;
 call symput("pertot&i",trim(left(pertot)));
            call symput("errtot&i",trim(left(errtot)));
 end;
 run;
 %let last = &j;
%end;
%let pertot = 0;
%let errtot = 0;
%do i=1 %to &num;
 %let pertot = %sysevalf(&pertot+&&pertot&i);
 %let errtot = %sysevalf(&errtot+&&errtot&i);
%end;
%put test = &test;
%put pertot = &pertot;
%put errtot = &errtot;
%mend;

%d(F,5,6 14 30 54 100,bm logbm age logage ht loght); * err = 10767;
%d(M,5,6 14 25 55 100,bm logbm age logage ht loght); * err = 22488;

proc sort data=males; by age; run;
proc means data=males noprint;
 by age;
 var bm logbm bmr ht loght;
 output out=m1 mean=;
run;
data m2;
 set m1;
 logage = log(1+age);
 if (age<=5) then fit = 13.19*bm + 270.2 *logbm - 18.34*age + 131.3*logage -
208.5 ;
 if (age>=6 and age<=13) then fit = 10.21*bm + 260.2 *logbm + 13.04*age -
205.7*logage + 333.4 ;
 if (age>=14 and age<=24) then fit = 0.207*bm + 1078. *logbm + 115.1*age -
2794.*logage + 3360.6;
 if (age>=25 and age<=54) then fit = 2.845*bm + 729.6 *logbm + 3.181*age -
191.6*logage - 1067. ;
 if (age>=55) then fit = 9.291*bm + 264.8 *logbm - 5.288*age + 181.5*logage -
705.9 ;
 if (fit<50) then fit=50;
 if (fit>3000) then fit=3000;
 if (age<=5) then fit2 = 17.61*bm + 106.3 *logbm - 17.93*age + 87.37*logage -
368.9*ht + 676.3 *loght + 607.6;
 if (age>=6 and age<=13) then fit2 = 12.64*bm + 149.3 *logbm + 30.91*age -
417.0*logage - 1498.*ht + 2151.5*loght + 2344.9;
 if (age>=14 and age<=24) then fit2 = .0309*bm + 1098.6*logbm + 114.3*age -
2777.*logage + 31.45*ht - 101.2 *loght + 3250.7;
 if (age>=25 and age<=54) then fit2 = 4.692*bm + 481.5 *logbm + 2.422*age -
136.3*logage + 1590.*ht - 2014. *loght - 1961.3;
```

```
 if (age>=55) then fit2 = 12.60*bm - 108.4 *logbm - 5.151*age + 170.6*logage
- 927.2*ht + 2405. *loght + 982.6;
 if (fit<50) then fit=50;
 if (fit>3000) then fit=3000;
 if (age<=2) then old = 0.249*bm - 0.127 ;
 if (age>=3 and age<=9) then old = 0.095*bm + 2.110 ;
 if (age>=10 and age<=17) then old = 0.074*bm + 2.754 ;
 if (age>=18 and age<=29) then old = 0.063*bm + 2.896 ;
 if (age>=30 and age<=59) then old = 0.048*bm + 3.653 ;
 if (age>=60) then old = 0.049*bm + 2.459 ;
 old = 238.845 * old;
 if (old<144) then old=144;
 if (old>2880) then old=2880;
run;
symbol1 color=black;
symbol2 color=red;
symbol3 color=blue;
symbol4 color=green;
title "mean bmr and old and new fits - males";
title2 "data=black, old fit=red, new fit=blue, new with ht=green";
proc gplot data=m2;
 plot old*age=2 fit*age=3 bmr*age=1 fit2*age=4 /overlay;
run; quit;
proc gplot data=m2(where=(age>=48 and age<=63));
 plot old*age=2 fit*age=3 bmr*age=1 fit2*age=4 /overlay;
run; quit;



proc sort data=females; by age; run;
proc means data=females noprint;
 by age;
 var bm logbm bmr ht loght;
 output out=f1 mean=;
run;
data f2;
 set f1;
 logage = log(1+age);
 if (age<=5) then fit = 11.94*bm + 261.5 *logbm - 22.31*age + 120.9*logage -
183.6;
 if (age>=6 and age<=13) then fit = 5.296*bm + 409.1 *logbm + 40.37*age -
524.9*logage + 392.7;
 if (age>=14 and age<=29) then fit = 0.968*bm + 676.9 *logbm + 40.89*age -
1002.*logage + 772.7;
 if (age>=30 and age<=53) then fit = 4.935*bm + 355.4 *logbm + 16.28*age -
896.0*logage + 2225.;
 if (age>=54) then fit = 2.254*bm + 445.9 *logbm + 5.464*age - 489.9*logage +
944.2;
 if (fit<50) then fit=50;
 if (fit>3000) then fit=3000;
 if (age<=5) then fit2 = 21.78*bm - 16.26 *logbm - 9.014*age + 39.09 *logage
- 942.8 *ht + 1259.9*loght + 1443.0;
 if (age>=6 and age<=13) then fit2 = 7.540*bm + 262.8 *logbm + 43.41*age -
604.3 *logage - 338.0 *ht + 758.7 *loght + 1209.3;
 if (age>=14 and age<=29) then fit2 = 4.194*bm + 391.6 *logbm + 41.38*age -
1010.3*logage + 152.5 *ht + 433.1 *loght + 1298.2;
```

```
 if (age>=30 and age<=53) then fit2 = 6.239*bm + 208.5 *logbm + 14.38*age -
803.3 *logage + 2854.4*ht - 4066. *loght - 180.9;
 if (age>=54) then fit2 = 3.840*bm + 284.9 *logbm + 4.510*age - 400.1 *logage
+ 1782.8*ht - 2274. *loght - 588.6;
 if (fit<50) then fit=50;
 if (fit>3000) then fit=3000;
 if (age<=2) then old = 0.244*bm - 0.130 ;
 if (age>=3 and age<=9) then old = 0.085*bm + 2.033 ;
 if (age>=10 and age<=17) then old = 0.056*bm + 2.898 ;
 if (age>=18 and age<=29) then old = 0.062*bm + 2.036 ;
 if (age>=30 and age<=59) then old = 0.034*bm + 3.538 ;
 if (age>=60) then old = 0.038*bm + 2.755 ;
 old = 238.845 * old;
 if (old<144) then old=144;
 if (old>2880) then old=2880;
run;
symbol1 color=black;
symbol2 color=red;
symbol3 color=blue;
symbol4 color=green;
title "mean bmr and old and new fits - females";
title2 "data=black, old fit=red, new fit=blue, new with ht=green";
proc gplot data=f2;
 plot bmr*age=1 old*age=2 fit*age=3 fit2*age=4 /overlay;
run; quit;
proc gplot data=f2(where=(age>=48 and age <=63));
 plot bmr*age=1 old*age=2 fit*age=3 fit2*age=4 /overlay;
run; quit;

data mall;
 set males;
 z = rannor(0);
 if (age<=5) then fit = 13.19*bm + 270.2 *logbm - 18.34*age + 131.3*logage -
208.5 + 69.10*z;
 if (age>=6 and age<=13) then fit = 10.21*bm + 260.2 *logbm + 13.04*age -
205.7*logage + 333.4 + 115.3*z;
 if (age>=14 and age<=29) then fit = 0.207*bm + 1078. *logbm + 115.1*age -
2794.*logage + 3360.6 + 161.1*z;
 if (age>=30 and age<=53) then fit = 2.845*bm + 729.6 *logbm + 3.181*age -
191.6*logage - 1067. + 178.2*z;
 if (age>=54) then fit = 9.291*bm + 264.8 *logbm - 5.288*age + 181.5*logage -
705.9 + 163.6*z;
 if (fit<50) then fit=50;
 if (fit>3000) then fit=3000;
 if (age<=5) then fit2 = 11.59*bm + 215.6 *logbm - 29.69*age + 112.9*logage +
367.1*ht - 332.7 + 68.93*z;
 if (age>=6 and age<=13) then fit2 = 10.42*bm + 239.4 *logbm + 11.87*age -
200.3*logage + 42.18*ht + 339.8 + 115.3*z;
 if (age>=14 and age<=24) then fit2 = 0.103*bm + 1094. *logbm + 114.4*age -
2781.*logage - 28.7*ht + 3322.1 + 161.1*z;
 if (age>=25 and age<=54) then fit2 = 5.022*bm + 457.5 *logbm + 2.370*age -
134.5*logage + 405.3*ht - 939.6 + 176.7*z;
 if (age>=55) then fit2 = 11.78*bm - 44.62 *logbm - 3.177*age + 39.95*logage
+ 490.8*ht + 50.55 + 160.9*z;
 if (fit2<50) then fit2=50;
 if (fit2>3000) then fit2=3000;
```

```
 if (age<=5) then fit3 = 17.61*bm + 106.3 *logbm - 17.93*age + 87.37*logage -
368.9*ht + 676.3*loght + 607.6 + 68.60*z;
 if (age>=6 and age<=13) then fit3 = 12.64*bm + 149.3 *logbm + 30.92*age -
417.0*logage - 1498.*ht + 2151.*loght + 2344.9 + 115.0*z;
 if (age>=14 and age<=24) then fit3 = .0309*bm + 1098.6*logbm + 114.3*age -
2777.*logage + 31.45*ht - 101.2*loght + 3250.7 + 161.1*z;
 if (age>=25 and age<=54) then fit3 = 4.692*bm + 481.5 *logbm + 2.422*age -
136.3*logage + 1590.*ht - 2014.*loght - 1961.3 + 176.6*z;
 if (age>=55) then fit3 = 12.67*bm - 113.9 *logbm - 3.228*age + 38.95*logage
- 962.2*ht + 2466.*loght + 1453.5 + 160.9*z;
 if (fit3<50) then fit3=50;
 if (fit3>3000) then fit3=3000;
 if (ht=.) then fit3=.;
 if (age<=2) then old = 0.249*bm - 0.127 + 0.29*z;
 if (age>=3 and age<=9) then old = 0.095*bm + 2.110 + 0.28*z;
 if (age>=10 and age<=17) then old = 0.074*bm + 2.754 + 0.44*z;
 if (age>=18 and age<=29) then old = 0.063*bm + 2.896 + 0.64*z;
 if (age>=30 and age<=59) then old = 0.048*bm + 3.653 + 0.70*z;
 if (age>=60) then old = 0.049*bm + 2.459 + 0.69*z;
 old = 238.845 * old;
 if (old<144) then old=144;
 if (old>2880) then old=2880;
 err = BMR-fit;
 err2 = BMR-fit2;
 err3 = BMR-fit3;
 err0 = BMR-old;
run;
axis1 order = 0 to 3000 by 1000;
title "fitted bmr - all males";
proc gplot data=mall;
 plot fit*age;
run; quit;
title "fitted bmr with height - all males";
proc gplot data=mall;
 plot fit2*age;
run; quit;
title "fitted bmr with ht and loght - all males";
proc gplot data=mall;
 plot fit3*age=3;
run; quit;
title "APEX fit for bmr - all males";
proc gplot data=mall;
 plot old*age /vaxis=axis1;
run; quit;
title "error statistics - males";
proc means data=mall n mean std var min max;
 var bmr err0 err err2 err3;
run;
proc sort data=mall; by age; run;
proc means data=mall noprint;
 by age;
 var bmr fit fit2 fit3 old err err2 err3 err0;
 output out=mstats mean=;
run;
symbol1 color=black;
symbol2 color=red;
```

```
symbol3 color=blue;
title "mean bmr and old and new fits - males";
title2 "data=black, old fit=red, new fit=blue";
proc gplot data=mstats;
 plot old*age=2 fit*age=3 bmr*age=1 /overlap;
run; quit;

data fall;
 set females;
 z = rannor(0);
 if (age<=5) then fit = 11.94*bm + 261.3 *logbm - 22.14*age + 120.4*logage -
182.9 + 64.62*z;
 if (age>=6 and age<=13) then fit = 5.296*bm + 409.1 *logbm + 40.37*age -
524.9*logage + 392.7 + 99.43*z;
 if (age>=14 and age<=29) then fit = 1.004*bm + 674.4 *logbm + 41.11*age -
1007.*logage + 790.6 + 143.2*z;
 if (age>=30 and age<=53) then fit = 4.935*bm + 355.4 *logbm + 16.29*age -
896.0*logage + 2225.3 + 145.3*z;
 if (age>=54) then fit = 2.699*bm + 415.7 *logbm + 8.701*age - 711.6*logage +
1756.8 + 124.6*z;
 if (fit<50) then fit=50;
 if (fit>3000) then fit=3000;
 if (age<=5) then fit2 = 11.09*bm + 175.3 *logbm - 35.26*age + 98.50 *logage
+ 449.0*ht - 304.3 + 63.23*z;
 if (age>=6 and age<=13) then fit2 = 6.494*bm + 304.9 *logbm + 31.99*age -
483.8 *logage + 209.0*ht + 411.8 + 98.89*z;
 if (age>=14 and age<=29) then fit2 = 4.107*bm + 396.9 *logbm + 41.32*age -
1009.3*logage + 423.2*ht + 1049.9 + 141.1*z;
 if (age>=30 and age<=53) then fit2 = 6.969*bm + 155.6 *logbm + 14.74*age -
815.2 *logage + 316.4*ht + 2175.2 + 144.0*z;
 if (age>=54) then fit2 = 5.038*bm + 198.6 *logbm + 7.630*age - 610.7 *logage
+ 346.1*ht + 1602.5 + 122.6*z;
 if (fit2<50) then fit2=50;
 if (fit2>3000) then fit2=3000;
 if (age<=5) then fit3 = 21.78*bm - 16.26 *logbm - 9.014*age + 39.09 *logage
- 942.8 *ht + 1259.9*loght + 1443.0 + 61.89*z;
 if (age>=6 and age<=13) then fit3 = 7.540*bm + 262.8 *logbm + 43.41*age -
604.3 *logage - 338.0 *ht + 758.7 *loght + 1209.3 + 98.85*z;
 if (age>=14 and age<=29) then fit3 = 4.194*bm + 391.6 *logbm + 41.38*age -
1010.3*logage + 152.5 *ht + 423.1 *loght + 1298.2 + 141.1*z;
 if (age>=30 and age<=53) then fit3 = 6.239*bm + 208.5 *logbm + 14.38*age -
803.3 *logage + 2854.4*ht - 4066. *loght - 180.9 + 143.9*z;
 if (age>=54) then fit3 = 4.506*bm + 236.4 *logbm + 7.564*age - 605.8 *logage
+ 1489.9*ht - 1796.6*loght + 475.8 + 122.6*z;
 if (fit3<50) then fit3=50;
 if (fit3>3000) then fit3=3000;
 if (ht=.) then fit3=.;
 if (age<=2) then old = 0.244*bm - 0.130 + 0.25*z;
 if (age>=3 and age<=9) then old = 0.085*bm + 2.033 + 0.29*z;
 if (age>=10 and age<=17) then old = 0.056*bm + 2.898 + 0.47*z;
 if (age>=18 and age<=29) then old = 0.062*bm + 2.036 + 0.50*z;
 if (age>=30 and age<=59) then old = 0.034*bm + 3.538 + 0.47*z;
 if (age>=60) then old = 0.038*bm + 2.755 + 0.45*z;
 old = 238.845 * old;
 if (old<144) then old=144;
 if (old>2880) then old=2880;
```

```
 err = BMR-fit;
 err2 = BMR-fit2;
 err3 = BMR-fit3;
 err0 = BMR-old;
run;
title "fitted bmr - all females";
proc gplot data=fall;
 plot fit*age;
run; quit;
title "fitted bmr with height - all females";
proc gplot data=fall;
 plot fit2*age;
run; quit;
title "fitted bmr with ht and loght - all females";
proc gplot data=fall;
 plot fit3*age=3;
run; quit;
axis1 order = 0 to 3000 by 1000;
title 'BMR - all males';
proc gplot data=mall;
 plot bmr*age /vaxis=axis1;
run; quit;
title 'BMR - all females';
proc gplot data=fall;
 plot bmr*age /vaxis=axis1;
run; quit;
proc means data=fall n mean std var min max;
 var bmr err0 err err2 err3;
run;
proc sort data=fall; by age; run;
proc means data=fall noprint;
 by age;
 var bmr fit fit2 fit3 old err err2 err3 err0;
 output out=fstats mean=;
run;
symbol1 color=black;
symbol2 color=red;
symbol3 color=blue;
title "mean bmr and old and new fits - females";
title2 "data=black, old fit=red, new fit=blue";
proc gplot data=fstats;
 plot old*age=2 fit*age=3 bmr*age=1 /overlap;
run; quit;
proc means data=males(where=(ht NE .)) n mean std var; var bmr; run;
proc means data=females(where=(ht NE .)) n mean std var; var bmr; run;
```

```
* August 2, 2016 by WGG, based on program by Jonathan Cohen;

libname apex 'C:\main\APEX\WA342\task4\task4';


data adams4;
 set apex.adams4 end=eof;
 * The following four obs deleted by JEL email of 3/2/2016;
 if STUD = 2 and SUBJ = 32 and step = 1.0 then delete;
 if STUD = 2 and SUBJ = 38 and step = 1.0 then delete;
 if STUD = 20 and SUBJ = 8 and step = 5.0 then delete;
 if STUD = 30 and SUBJ = 114 and step = 0.1 then delete;
 if ve=. or ln_vo2=. or vo2m=. or gend=. or age=. then delete;
 * VO2 units are L/min;
 vo2 = exp(ln_vo2);
 * VO2m is personal maximum VO2 in L/min;
 retain sum1 0;
 sum1 = sum1 + ve;
 if (eof) then do;
 meanve= sum1/_N_;
 call symput ("mean_ve",trim(left(meanve)));
 end;
 * Macro variable mean_ve is used later in calculating r2 for ve;
 drop sum1 meanve;
 label vo2='VO2';
run;
proc sort data=adams4 out=sorted; by stud subj; run;
data persons;
 set sorted;
 by stud subj;
 retain vo2max nobs 0;
 keep stud subj nobs vo2m vo2max;
 if first.subj then do; nobs=0; vo2max=vo2m; end;
 nobs = nobs+1;
 if vo2max<vo2 then vo2max=vo2;
 if last.subj then output;
run;
proc freq data=persons; tables nobs; run;

data base;
 merge sorted persons;
 by stud subj;
 retain reset 0;
 invm = 1/vo2m;
 logm = log(vo2m);
 * f1 is fraction of personal maximum (unitless);
 f1 = vo2/vo2m;
 f2 = f1**2;
 f3 = f1**3;
 f4 = f1**4;
 f5 = f1**5;
 g1 = vo2/vo2max;
 g2 = f1**2;
```

```
 g3 = f1**3;
 g4 = f1**4;
 g5 = f1**5;
 * bmi is body mass index;
 bmi = bm/(ht/100)**2;
 ln_bmi = log(bmi);
 * ht is height in cm;
 ln_ht = log(ht);
 * bm is body mass in kg;
 ln_bm = log(bm);
 * age in full years - log uses age rounded up to prevent log(0);
 ln_age = log(1+age);
 id = _N_;
 * Gend=-1 are males, gend=1 are females;
run;

***********************************************;
*Box-cox analysis to assess y transformation. Run one model statement at a
time;
proc transreg data = base;
 * model boxcox(ve / lambda= -1 -0.5 -0.3333 -0.25 -0.2 -0.1666 -0.14286 -
0.125 -0.1111 -0.1 0 0.5 1 )= identity(ln_vo2); * -0.2;
 * model boxcox(ve / lambda= -1 -0.5 -0.3333 -0.25 -0.2 -0.1666 -0.14286 -
0.125 -0.1111 -0.1 0 0.5 1 )= identity(ln_vo2 f1); * -0.125;
 * model boxcox(ve / lambda= -1 -0.5 -0.3333 -0.25 -0.2 -0.1666 -0.14286 -
0.125 -0.1111 -0.1 0 0.5 1 )= identity(ln_vo2 f2); * -0.1;
 * model boxcox(ve / lambda= -1 -0.5 -0.3333 -0.25 -0.2 -0.1666 -0.14286 -
0.125 -0.1111 -0.1 0 0.5 1 )= identity(ln_vo2 f3); * 0;
 model boxcox(ve / lambda= -1 -0.5 -0.3333 -0.25 -0.2 -0.1666 -0.14286 -0.125
-0.1111 -0.1 0 0.5 1 )= identity(ln_vo2 f4); * 0;
 * model boxcox(ve / lambda= -1 -0.5 -0.3333 -0.25 -0.2 -0.1666 -0.14286 -
0.125 -0.1111 -0.1 0 0.5 1 )= identity(ln_vo2 f5); * 0;
run;

/* With just ln_vo2, the best transformation is lambda=-0.2. With higher
powers of vo2/vo2m included
 this shifts to 0, which is the log transform.
*/


%macro regr(power,x);
 data a;
 set base end=eof;
 if (&power>0) then y = ve**(-1/(&power));
      else y = log(ve);
 run;
 *calculate regression coefficients & include VIF;
 proc reg data=a noprint;
 model y = &x/ vif;
 output out=b
       p =predicted
       residual=residual
       rstudent=rstudent;
 run; quit;
 *remove studentized outliers;
 data c;
```

```
 set b;
 if rstudent = . then delete;
 if abs(rstudent) > 3 then delete;
 run;
 * Redo regression without outliers;
 proc reg data=c plots(maxpoints=6700);
      model y = &x/ vif;
      output out=d
 p =predicted2
 residual=residual2
 rstudent=rstudent2;
 run; quit;
 * Calculate and report r2 on the original variable ve;
 data e;
 set d end=eof;
 if (&power>0) then pred = 1/predicted2**(&power);
      else pred = exp(predicted2);
 retain sumb sum1 0;
 db = (ve-&mean_ve)**2;
 d1 = (ve-pred)**2;
 sumb = sumb + db;
 sum1 = sum1 + d1;
 if (eof) then do;
 vb = sumb / _N_;
       v1 = sum1 / _N_;
       stat1 = 1 - v1/vb;
       put "vars &x ";
 put "stats " _N_ sumb sum1 vb v1 stat1;
 end;
      keep stud subj ve vo2 ln_vo2 vo2m y f1 f2 f3 f4 f5 gend pred;
 run;
%mend regr;

%regr(2, ln_vo2)              * tr_r2 = 0.9479        ve_r2 = 0.7350;
%regr(3, ln_vo2)              * tr_r2 = 0.9566        ve_r2 = 0.8779;
%regr(4, ln_vo2)              * tr_r2 = 0.9563        ve_r2 = 0.8873;
%regr(5, ln_vo2)              * tr_r2 = 0.9544        ve_r2 = 0.8850;
%regr(6, ln_vo2)              * tr_r2 = 0.9523        ve_r2 = 0.8821;
%regr(0, ln_vo2)              * tr_r2 = 0.9341        ve_r2 = 0.8561;

%regr(4, ln_vo2)  * tr_r2 = 0.9563        ve_r2 = 0.8873;
%regr(4, ln_vo2 age)    * tr_r2 = 0.9581        ve_r2 = 0.8900;
%regr(4, ln_vo2 gend)    * tr_r2 = 0.9578       ve_r2 = 0.8923;
%regr(4, ln_vo2 ht)     * tr_r2 = 0.9596        ve_r2 = 0.8938;
%regr(4, ln_vo2 vo2m)   * tr_r2 = 0.9715        ve_r2 = 0.9213;
%regr(4, ln_vo2 f1)           * tr_r2 = 0.9721        ve_r2 = 0.9378;
%regr(4, ln_vo2 f2)           * tr_r2 = 0.9712        ve_r2 = 0.9347;

%regr(5, ln_vo2 f1)     * tr_r2 = 0.9730        ve_r2 = 0.9402;
%regr(5, ln_vo2 f2)     * tr_r2 = 0.9729        ve_r2 = 0.9420;
%regr(5, ln_vo2 f3)     * tr_r2 = 0.9723        ve_r2 = 0.9402;
%regr(6, ln_vo2 f1)     * tr_r2 = 0.9730        ve_r2 = 0.9397;
%regr(6, ln_vo2 f2)     * tr_r2 = 0.9734  ve_r2 = 0.9445;
%regr(6, ln_vo2 f3)           * tr_r2 = 0.9731        ve_r2 = 0.9442;
%regr(6, ln_vo2 f4)     * tr_r2 = 0.9723        ve_r2 = 0.9427;
%regr(0, ln_vo2 f1)     * tr_r2 = 0.9662        ve_r2 = 0.9244;
```

```
%regr(0, ln_vo2 f2)      * tr_r2 = 0.9714        ve_r2 = 0.9411;
%regr(0, ln_vo2 f3)      * tr_r2 = 0.9724        ve_r2 = 0.9466;
%regr(0, ln_vo2 f4)      * tr_r2 = 0.9719        ve_r2 = 0.9481; * best;
%regr(0, ln_vo2 f5)      * tr_r2 = 0.9711 ve_r2 = 0.9479;


%regr(0, ln_vo2 f4 age)        * tr_r2 = 0.9720        ve_r2 = 0.9477;
%regr(0, ln_vo2 f4 gend)       * tr_r2 = 0.9721        ve_r2 = 0.9483;
%regr(0, ln_vo2 f4 ht)   * tr_r2 = 0.9723        ve_r2 = 0.9481;
%regr(0, ln_vo2 f4 gend age ht)      * tr_r2 = 0.9726 ve_r2 = 0.9477;

* For comparison, repeat the near-optimal regression using vo2max instead of
vo2m;
%regr(0, ln_vo2 g4) * ve_r2 = 0.9481;

/* %regr(0, ln_vo2 f4) seems to be the best choice. While very high powers
(11+) of 1/ve
 give marginally better r2, the log is a more usual choice, especially since
the primary
 independent variable (vo2) is also log transformed.

 Note: ve_r2 is based on the no-outlier data set (3 studentized residuals);
 On full Adams data set with (0, ln_vo2 f4), 6636 obs, r2 = 0.9463, which can
be
 checked by running %stats(adams4) below.;


Macro %stats examines the optimal choice, examining the effects of truncating
outliers
 on the predicted points. It does not seem to make much difference whether
the N(0,1)
 is truncated or not, or whether the generated ve values are truncated or
not. Note that
 %stats may be re-run several times, and the predicted values will change
because new
 random numbers are being drawn.
*/


%macro stats(ds);
proc sort data=&ds out=s; by stud subj; run;
data cloud;
 set s end=eof;
 by stud subj;
 retain ss vv v1 v1b v2 v2b q1 q1b t1 t1b 0;
 ve0 = min(ve,220);
 z = rannor(0);
 retain zb 0;
 if first.subj then zb = rannor(0);
 p1 = exp(3.29821+0.79351*ln_vo2+0.53845*f4);
 p1b = min(max(p1,4),220);
 ve1 = exp(3.29821+0.79351*ln_vo2+0.53845*f4+0.12529*z);
 ve1b = min(max(ve1,4),220);
 ve2 = exp(3.300+0.8128*ln_vo2+0.5126*f4+0.09866*zb+0.07852*z);
 ve2b = min(max(ve2,4),220);
 old = 1/(0.163-0.0816*ln_vo2-0.000342*age-0.00348*gend+0.000233*ht)**2;
```

```
 oldb = min(max(old,4),220);
 ss = ss + ve**2;
 q1 = q1 + (p1-ve)**2;
 q1b = q1b + (p1b-ve)**2;
 t1 = t1 + (old-ve)**2;
 t1b = t1b + (oldb-ve)**2;
 vv = vv + (ve-&mean_ve)**2;
 v1 = v1 + (ve1-&mean_ve)**2;
 v1b = v1b + (ve1b-&mean_ve)**2;
 v2 = v2 + (ve2-&mean_ve)**2;
 v2b = v2b + (ve2b-&mean_ve)**2;
 if (eof) then do;
 put "data set = &ds";
 put ss vv v1 v1b v2 v2b;
       qq1 = 1-q1/vv;
       qq1b = 1-q1b/vv;
       tt1 = 1-t1/vv;
       tt1b = 1-t1b/vv;
       put q1 q1b qq1 qq1b tt1 tt1b;
 end;
run;
%mend;

%stats(base)
%stats(e)

axis1 order = 0 to 220 by 20;
proc gplot data=cloud;
      plot ve0*vo2 /VAXIS=axis1;
      plot ve2*vo2 /VAXIS=axis1;
run;quit;

proc means data=cloud N min mean median std max;
 var ve ve1 ve2 old;
run;

proc mixed data=e covtest plots(maxpoints=6700);
      class stud subj;
      model y = ln_vo2 f4 /solution ddfm=kr;
      random subj(stud)/ solution ;
      title 'data= random statement & ddfm=kr';
      ods output covParms=mixedcovm_old;
      ods output solutionF=solutions_old;
run;
```

APPENDIX I

CONSOLIDATED HUMAN ACTIVITY DATABASE (CHAD) DATA

A total of 24 Consolidated Human Activity Database (CHAD) studies were included in CHAD as of November 2015, with 179,912 diary-days entered. The geographic coverages range from specific cities to collections of metropolitan areas to the entire US, and the respondents tend to be adults but some studies include (or are limited to) children. CHAD contains human activity data from these studies, coded into a harmonized set of location and activity codes. Note, however, that the data collected in the original studies differed in level of detail in terms of activity, location, and time resolution. In addition, the translation of the original study data into CHAD format was performed by different individuals or groups. Therefore, the CHAD data themselves will vary in specificity and resolution across the studies. One of the goals of this manual is to provide any user with enough information to assess each study within CHAD for appropriateness for their application. An overview of the studies is provided in Table I-1 below.

Table I-1. Overview of Activity Studies Included in CHAD-Master (as of November 2015)

Study Name	Geographic Coverage	Dates (as incorporated into CHAD)	Respondent Ages (years; as incorporated into CHAD)	Data Gathering	Diary-Days (as incorporated into CHAD)	Study References
Baltimore Retirement Home Study (BAL)	Baltimore County, MD	01–02/1997 07–08/1998	≥65	daily recall data collected by study staff over a 3-week period	391	Williams et al., 2000
American Time Use Survey, Bureau of Labor Statistics (BLS)	Whole US	2003–2011	≥15	24-hour recall data collected by telephone interview combining structured questions and conversational interviewing	124,517	BLS, 2014
California Activity Pattern Studies (CAA, CAC, CAY)	California	CAA and CAY: 10/1987–09/1988 CAC: 04/1989–02/1990	CAA: 18–94 CAY: 12–17 CAC: ≤11	24-hour recall data collected by telephone interviews with structured questions	CAA: 1,579 CAY: 183 CAC: 1,200	Wiley et al., 1991a; 1991b
Cincinnati Activity Patterns Study (CIN)	Cincinnati, OH	08–09/1985	≤86	activity diary and background questionnaire	2,614	Johnson, 1989
Detroit Exposure and Aerosol Research Study (DEA)	Detroit, MI	06/2004–10/2007	≥18	activities recorded via free-form entry, while location data were structured	340	Williams et al., 2008
Denver, Colorado Personal Exposure Study (DEN)	Denver, CO	11/1982–02/1983	18–70	activity diary and background questionnaire	805	Johnson, 1984; Johnson et al., 1986
EPA Longitudinal Studies (EPA)	Respondents residing in Central NC (Raleigh, Durham, Chapel Hill)	1999–2000, 2002, 2006–2008, 2012–2013	0, 35–67	paper diary; free-from questionnaire	1,786	Isaacs et al., 2012
Population Study of Income Dynamics PSID I, II, III (ISR)	Whole US	I: 02–12/1997I I: 2002–2003 III:09/2007–05/2005	I: ≤12 II and II: <18	interviews; time diaries	I: 5,616 II: 4,997 III: 2,741	Alion Science and Technology, 2012; University of Michigan, 2014

Study Name	Geographic Coverage	Dates (as incorporated into CHAD)	Respondent Ages (years; as incorporated into CHAD)	Data Gathering	Diary-Days (as incorporated into CHAD)	Study References
Los Angeles Ozone Exposure Study: Elementary School/High School (LAE/LAH)	Los Angeles, CA	Fall/1989, Fall/1990	10–17	real-time diaries	94	Roth Associates, 1988; Spier et al., 1992
North Carolina State University Study (NCS)	Mostly NC, 9 other states also included	09–10/2013, 09–10/2014	22–58	diaries recorded in real time	662	Hill, 2014
National Human Activity Pattern Study (NHAPS): Air/Water (NHA/NHW)	48 states	09/1992–10/1994	≤93	telephone interview and questionnaire	NHA: 4,723 NHW: 4,663	Klepeis et al., 1995; Tsang and Klepeis, 1996
National-scale Activity Study (NSA)	7 metropolitan areas	06–09/2009	35–92	recall activity diary questionnaire	6,862	Knowledge Networks, 2009
RTI Ozone Averting Behavior Study (OAB)	35 metropolitan areas	07–09/2002, 08/2003	2–12	no information provided at this time	2,907	Mansfield et al., 2009
RTP Particulate Matter Panel Study (RTP)	Wake and Orange Counties, NC	06–11/2000, 01–05/2001	55–85	diaries recorded in real time	998	Williams et al., 2001; 2003a,b
Seattle Study (SEA)	Seattle, WA	10/1999–05/2001	6–91	diaries recorded in real time	1,692	Liu et al., 2003
Study of Use of Products and Exposure-related Behaviors (SUP)	California	06/2006–03/2010	≤88	24-hour recall data, collected by phone interview	9,446	Bennett et al., 2012
Valdez Air Health Study (VAL)	Valdez, AK	04–05/1990, 08/1990, 02–03/1991	11–71	information not provided	397	Goldstein et al., 1992
Washington, DC Study (WAS)	Washington, DC	11/1982–02/1983	18–71	activity diary and background questionnaire	699	Hartwell et al., 1984; Johnson et al., 1986; Settergren et al., 1984

REFERENCES

Alion Science and Technology. (2012). PSID Integration into CHAD (a description from Alion on integrating ISR into CHAD).

Bennett DH, Teague CH, Lee K, Cassady DL, Ritz B, and Hertz-Picciotto I. (2012). Passive sampling methods to determine household and personal care product use. Journal of Exposure Science and Environmental Epidemiology 22(2): 148–160.

BLS (Bureau of Labor Statistics). (2014). American Time Use Survey User's Guide: Understanding ATUS 2003 to 2013; Bureau of Labor Statistics, Washington, DC; December 2014. Available at: http://www.bls.gov/tus/atususersguide.pdf.

Goldstein B, Tardiff R, Hoffnagle G, and Kester R. (1992). Valdez Air Health Study: Summary Report. Prepared for Alyeska Pipeline Service Company, Anchorage, AK.

Hill Z., 2014. Development and Evaluation of Human Longitudinal Time-Location-Activity Data. Available at: http://www.lib.ncsu.edu/resolver/1840.4/8287.

Hartwell TD, Clayton CA, Michie RM, Whitmore RW, Zelon HS, Jones SM, and Whitehurst DA. (1984). Study of Carbon Monoxide Exposure of Residents of Washington, D.C. and Denver, Colorado. Prepared for the U.S. Environmental Protection Agency. Research Triangle Park, NC.

Isaacs K, McCurdy T, Glen G, Nysewander M, Errickson A, Forbes S, Graham S, McCurdy L, Smith L, Tulve N, and Vallero, D. (2012). Statistical properties of longitudinal time-activity data for use in human exposure modeling. Journal of Exposure Science and Environmental Epidemiology 23(3): 328-336.

Johnson, T. (1984). Study of Personal Exposure to Carbon Monoxide in Denver, Colorado. Prepared for U.S. Environmental Protection Agency, Environmental Monitoring Systems Laboratory, Research Triangle Park, NC.

Johnson, T, Capel J, and Wijnberg L. (1986). Selected Data Analyses Relating to Studies of Personal Carbon Monoxide Exposure in Denver and Washington, DC. Prepared for U.S. Environmental Protection Agency, Environmental Monitoring Systems Laboratory, Research Triangle Park, NC.

Johnson, T. (1989). Human Activity Patterns in Cincinnati, Ohio. Final Report. Prepared for Electric Power Research Institute, Health Studies Program, Palo Alto, CA.

Klepeis N, Tsang A, and Behar J. (1995). Analysis of the National Human Activity Pattern Survey (NHAPS) Respondents from a Standpoint of Exposure Assessment. Final Report. Prepared for U.S. Environmental Protection Agency, National Exposure Research Laboratory, Las Vegas, NV.

Knowledge Networks. (2009). Field Report: National Scale Activity Survey (NSAS). Conducted for Research Triangle Institute. Submitted to Carol Mansfield November 13, 2009.

Liu L-JS, Box M, Kalman D, Kaufman J, Koenig J, Larson T, Lumley T, Sheppard L, and Wallace L. 2003. Exposure assessment of particulate matter for susceptible populations in Seattle. Environ Health Perspect 111: 909–918.

Mansfield C, Houtven GV, Johnson F R, and Yang J-C. (2009). Environmental Risks and Behavior: Do children spend less time outdoors when ozone pollution is high? ASSA annual meeting, January 5, 2009. Update of Houtven et al. (2003) using the OAB CHAD data set, and related to Mansfield et al. (2006).

Roth Associates. (1988). LA_part1 and LA_part2 (A Study of Activity Patterns Among a Group of Los Angeles Asthmatics). Electric Power Research Institute

Settergren SK, Hartwell TD, and Clayton CA. (1984). Study of Carbon Monoxide Exposure of Residents of Washington, DC.: Additional Analyses. Prepared for U.S. Environmental Protection Agency, Environmental Monitoring Systems Laboratory, Research Triangle Park, NC.

Spier C, Little D, Trim S, Johnson T, Linn W, and Hackney J. (1992). Activity Patterns in Elementary and High School Students Exposed to Oxidant Pollution. Journal of Exposure Analysis and Environmental Epidemiology 2: 277–293.

Tsang AM and Klepeis NE. (1996). Descriptive Statistics Tables from a Detailed Analysis of the National Human Activity Pattern Survey (NHAPS) Data, U.S. Environmental Protection Agency, Washington, D.C.

University of Michigan. (2014). The Panel Study of Income Dynamics. http://psidonline.isr.umich.edu/Studies.aspx.

Wiley J, Robinson J, Piazza T, Garrett K, Cirksena K, Cheng Y, and Martin G. (1991a). Activity Patterns of California Residents. Final Report. Prepared for California Air Resources Board, Research Division, Sacramento, CA.

Wiley J, Robinson J, Cheng Y, Piazza T, Stork L, and Pladsen K. (1991b). Study of Children's Activity Patterns. Final Report under contract no A733-149. Prepared for California Air Resources Board, Research Division, Sacramento, CA.

Williams, R, Suggs, J, Creason, J, Rodes, C, Lawless, P, Kwok, R, Zweidinger, R, and Sheldon, L. (2000). The 1998 Baltimore particulate matter epidemiology-exposure study: Part 2. Personal exposure associated with an elderly population. J Expo Anal Environ Epidemiol. 10(6): 533–543.

Williams RW, Wallace LA, Suggs JC, Evans EG, Creason JP, Highsmith VR, Sheldon LS, Rea AW, Vette AF, Zweidinger RB, Leovic KW, Norris GA, Landis MS, HowardReed C, Stevens C, Conner TL, Rodes CE, Lawless PA, Thornburg J, Liu LS, Kalman D, Kaufman J, Koenig JQ, Larson TL, Lumley T, Sheppard L, Brown K, Suh H, Wheeler A, Gold D, Koutrakis P, and Lippmann M. (2001). Preliminary particulate matter mass concentrations associated with longitudinal panel studies: assessing human exposures of high risk subpopulations to particulate matter. Office of Research and Development. United States Environmental Protection Agency. EPA/600/R-01/086.

Williams R, Suggs J, Rea A, Leovic K, Vette A, Croghan C, Sheldon L, Rodes C, Thornburg J, Ejire A, Herbst M, and Sanders W. (2003a). The Research Triangle Park particulate matter panel study: PM mass concentration relationships. Atmospheric Environment 37 (38): 5349–5363.

Williams R, Suggs J, Rea A, Sheldon L, Rodes C, and Thornburg J. (2003b). The Research Triangle Park particulate matter panel study: Modeling ambient source contribution to personal and residential PM mass concentrations. Atmospheric Environment 37 (36): 5365–5378.

Williams R, Rea A, Vette A, Croghan C, Whitaker D, Stevens C, McDow A, Fortmann R, Sheldon L, Wilson H, Thornburg J, Phillips M, Lawless P, Rodes C, and Daughtrey H. (2008). The design and field implementation of the Detroit Exposure and Aerosol Research Study. Journal of Exposure Science and Environmental Epidemiology 19: 643–659.

APPENDIX J

DETAILED EXPOSURE AND RISK RESULTS

Table J-1. APEX estimates for percent of children and adults with asthma in Fall River study area, 2011.

Percent of children with asthma at elevated ventilation having exposures at or above 5-minute benchmark concentrations							
		number of days per year					
benchmark		at least 1	at least 2	at least 3	at least 4	at least 5	at least 6
100 ppb		32.74	12.17	5.49	2.55	1.31	0.62
200 ppb		0.24	0	0	0	0	0
300 ppb		0	0	0	0	0	0
400 ppb		0	0	0	0	0	0
Percent of children with asthma estimated to experience at least one day with an increase in sRaw ≥ 100%							
		number of days per year					
E-R Function	sRaw	at least 1	at least 2	at least 3	at least 4	at least 5	at least 6
LB	100%	0.41	0.14	0.05	0	0	0
	200%	0	0	0	0	0	0
MEAN	100%	1.43	0.71	0.47	0.38	0.27	0.22
	200%	0.25	0.11	0.05	0	0	0
UB	100%	3.68	2.50	1.95	1.57	1.26	1.13
	200%	1.48	0.99	0.77	0.60	0.52	0.52
Percent of adults with asthma at elevated ventilation having exposures at or above 5-minute benchmark concentrations							
		number of days per year					
benchmark		at least 1	at least 2	at least 3	at least 4	at least 5	at least 6
100 ppb		5.08	0.44	0.05	0	0	0
200 ppb		0.02	0	0	0	0	0
300 ppb		0	0	0	0	0	0
400 ppb		0	0	0	0	0	0
Percent of adults with asthma estimated to experience at least one day with an increase in sRaw ≥ 100%							
		number of days per year					
E-R Function	sRaw	at least 1	at least 2	at least 3	at least 4	at least 5	at least 6
LB	100%	0.06	0	0	0	0	0
	200%	0	0	0	0	0	0
MEAN	100%	0.34	0.09	0.03	0	0	0
	200%	0.05	0	0	0	0	0
UB	100%	1.28	0.55	0.33	0.21	0.16	0.11
	200%	0.56	0.25	0.15	0.10	0.07	0.05

Table J-2. APEX estimates for percent of children and adults with asthma in Fall River study area, 2012.

Percent of children with asthma at elevated ventilation having exposures at or above 5-minute benchmark concentrations						
	number of days per year					
benchmark	at least 1	at least 2	at least 3	at least 4	at least 5	at least 6
100 ppb	13.21	2.76	0.56	0.12	0.03	0
200 ppb	0	0	0	0	0	0
300 ppb	0	0	0	0	0	0
400 ppb	0	0	0	0	0	0

Percent of children with asthma estimated to experience at least one day with an increase in sRaw ≥ 100%							
		number of days per year					
E-R Function	sRaw	at least 1	at least 2	at least 3	at least 4	at least 5	at least 6
LB	100%	0.14	0.03	0	0	0	0
	200%	0	0	0	0	0	0
MEAN	100%	0.77	0.44	0.25	0.14	0.08	0.08
	200%	0.14	0.03	0	0	0	0
UB	100%	2.55	1.76	1.29	1.04	0.91	0.77
	200%	1.10	0.74	0.55	0.44	0.38	0.30

Percent of adults with asthma at elevated ventilation having exposures at or above 5-minute benchmark concentrations						
	number of days per year					
benchmark	at least 1	at least 2	at least 3	at least 4	at least 5	at least 6
100 ppb	1.86	0.18	0	0	0	0
200 ppb	0	0	0	0	0	0
300 ppb	0	0	0	0	0	0
400 ppb	0	0	0	0	0	0

Percent of adults with asthma estimated to experience at least one day with an increase in sRaw ≥ 100%							
		number of days per year					
E-R Function	sRaw	at least 1	at least 2	at least 3	at least 4	at least 5	at least 6
LB	100%	0.02	0	0	0	0	0
	200%	0	0	0	0	0	0
MEAN	100%	0.17	0.04	0	0	0	0
	200%	0.01	0	0	0	0	0
UB	100%	0.88	0.38	0.22	0.15	0.10	0.08
	200%	0.39	0.16	0.10	0.07	0.05	0.03

Table J-3. APEX estimates for percent of children and adults with asthma in Fall River study area, 2013.

Percent of children with asthma at elevated ventilation having exposures at or above 5-minute benchmark concentrations						
	number of days per year					
benchmark	**at least 1**	**at least 2**	**at least 3**	**at least 4**	**at least 5**	**at least 6**
100 ppb	12.29	1.60	0.33	0.03	0	0
200 ppb	0	0	0	0	0	0
300 ppb	0	0	0	0	0	0
400 ppb	0	0	0	0	0	0

Percent of children with asthma estimated to experience at least one day with an increase in sRaw ≥ 100%							
		number of days per year					
E-R Function	**sRaw**	**at least 1**	**at least 2**	**at least 3**	**at least 4**	**at least 5**	**at least 6**
LB	100%	0.11	0	0	0	0	0
	200%	0	0	0	0	0	0
MEAN	100%	0.55	0.08	0.05	0.03	0.03	0
	200%	0.05	0	0	0	0	0
UB	100%	1.95	1.04	0.77	0.60	0.52	0.44
	200%	0.77	0.44	0.33	0.27	0.25	0.19

Percent of adults with asthma at elevated ventilation having exposures at or above 5-minute benchmark concentrations						
	number of days per year					
benchmark	**at least 1**	**at least 2**	**at least 3**	**at least 4**	**at least 5**	**at least 6**
100 ppb	1.32	0.07	0	0	0	0
200 ppb	0	0	0	0	0	0
300 ppb	0	0	0	0	0	0
400 ppb	0	0	0	0	0	0

Percent of adults with asthma estimated to experience at least one day with an increase in sRaw ≥ 100%							
		number of days per year					
E-R Function	**sRaw**	**at least 1**	**at least 2**	**at least 3**	**at least 4**	**at least 5**	**at least 6**
LB	100%	0	0	0	0	0	0
	200%	0	0	0	0	0	0
MEAN	100%	0.07	0	0	0	0	0
	200%	0	0	0	0	0	0
UB	100%	0.54	0.24	0.15	0.11	0.08	0.07
	200%	0.24	0.11	0.07	0.06	0.04	0.03

Table J-4. APEX estimates for percent of children and adults with asthma in Indianapolis study area, 2011.

Percent of children with asthma at elevated ventilation having exposures at or above 5-minute benchmark concentrations						
		number of days per year				
benchmark		at least 1	at least 2	at least 3	at least 4	at least 5	at least 6
100 ppb		27.03	7.97	2.45	0.97	0.41	0.10
200 ppb		1.04	0	0	0	0	0
300 ppb		0.83	0	0	0	0	0
400 ppb		0.31	0	0	0	0	0
Percent of children with asthma estimated to experience at least one day with an increase in sRaw ≥ 100%							
		number of days per year					
E-R Function	sRaw	at least 1	at least 2	at least 3	at least 4	at least 5	at least 6
LB	100%	0.57	0.19	0.11	0.07	0.06	0.05
	200%	0.07	0	0	0	0	0
MEAN	100%	1.48	0.77	0.57	0.44	0.38	0.32
	200%	0.36	0.17	0.11	0.09	0.06	0.06
UB	100%	3.80	2.60	2.11	1.83	1.61	1.47
	200%	1.61	1.13	0.94	0.80	0.72	0.66
Percent of adults with asthma at elevated ventilation having exposures at or above 5-minute benchmark concentrations							
		number of days per year					
benchmark		at least 1	at least 2	at least 3	at least 4	at least 5	at least 6
100 ppb		4.28	0.60	0.10	0	0	0
200 ppb		0.12	0	0	0	0	0
300 ppb		0.09	0	0	0	0	0
400 ppb		0.07	0	0	0	0	0
Percent of adults with asthma estimated to experience at least one day with an increase in sRaw ≥ 100%							
		number of days per year					
E-R Function	sRaw	at least 1	at least 2	at least 3	at least 4	at least 5	at least 6
LB	100%	0.11	0.03	0.01	0	0	0
	200%	0.01	0	0	0	0	0
MEAN	100%	0.44	0.17	0.10	0.06	0.04	0.03
	200%	0.10	0.03	0.02	0.01	0.01	0.00
UB	100%	1.59	0.90	0.64	0.49	0.39	0.33
	200%	0.72	0.42	0.31	0.23	0.20	0.17

Table J-5. APEX estimates for percent of children and adults with asthma in Indianapolis study area, 2012.

Percent of children with asthma at elevated ventilation having exposures at or above 5-minute benchmark concentrations						
		number of days per year				
benchmark		at least 1	at least 2	at least 3	at least 4	at least 5	at least 6
100 ppb		22.30	7.66	2.55	0.97	0.41	0.21
200 ppb		0	0	0	0	0	0
300 ppb		0	0	0	0	0	0
400 ppb		0	0	0	0	0	0
Percent of children with asthma estimated to experience at least one day with an increase in sRaw ≥ 100%							
		number of days per year					
E-R Function	sRaw	at least 1	at least 2	at least 3	at least 4	at least 5	at least 6
LB	100%	0.41	0.18	0.11	0.07	0.05	0.04
	200%	0.02	0	0	0	0	0
MEAN	100%	1.29	0.74	0.54	0.42	0.37	0.32
	200%	0.32	0.17	0.09	0.09	0.06	0.06
UB	100%	3.53	2.51	2.06	1.81	1.62	1.49
	200%	1.50	1.09	0.91	0.79	0.74	0.65
Percent of adults with asthma at elevated ventilation having exposures at or above 5-minute benchmark concentrations							
		number of days per year					
benchmark		at least 1	at least 2	at least 3	at least 4	at least 5	at least 6
100 ppb		3.78	0.55	0.14	0.05	0.03	0.02
200 ppb		0	0	0	0	0	0
300 ppb		0	0	0	0	0	0
400 ppb		0	0	0	0	0	0
Percent of adults with asthma estimated to experience at least one day with an increase in sRaw ≥ 100%							
		number of days per year					
E-R Function	sRaw	at least 1	at least 2	at least 3	at least 4	at least 5	at least 6
LB	100%	0.09	0.02	0.01	0	0	0
	200%	0	0	0	0	0	0
MEAN	100%	0.41	0.17	0.09	0.07	0.05	0.03
	200%	0.09	0.03	0.02	0.01	0.01	0
UB	100%	1.54	0.90	0.64	0.48	0.39	0.32
	200%	0.69	0.43	0.30	0.24	0.20	0.16

Table J-6. APEX estimates for percent of children and adults with asthma in Indianapolis study area, 2013.

Percent of children with asthma at elevated ventilation having exposures at or above 5-minute benchmark concentrations							
		number of days per year					
benchmark		at least 1	at least 2	at least 3	at least 4	at least 5	at least 6
100 ppb		17.95	4.69	1.73	0.31	0.14	0.07
200 ppb		0.93	0	0	0	0	0
300 ppb		0	0	0	0	0	0
400 ppb		0	0	0	0	0	0
Percent of children with asthma estimated to experience at least one day with an increase in sRaw ≥ 100%							
		number of days per year					
E-R Function	sRaw	at least 1	at least 2	at least 3	at least 4	at least 5	at least 6
LB	100%	0.35	0.14	0.08	0.06	0.04	0.03
	200%	0.02	0	0	0	0	0
MEAN	100%	1.12	0.64	0.46	0.37	0.32	0.29
	200%	0.27	0.13	0.08	0.07	0.06	0.06
UB	100%	3.23	2.27	1.91	1.65	1.48	1.36
	200%	1.39	0.99	0.84	0.72	0.67	0.62
Percent of adults with asthma at elevated ventilation having exposures at or above 5-minute benchmark concentrations							
		number of days per year					
benchmark		at least 1	at least 2	at least 3	at least 4	at least 5	at least 6
100 ppb		2.90	0.36	0.12	0.03	0	0
200 ppb		0.17	0	0	0	0	0
300 ppb		0	0	0	0	0	0
400 ppb		0	0	0	0	0	0
Percent of adults with asthma estimated to experience at least one day with an increase in sRaw ≥ 100%							
		number of days per year					
E-R Function	sRaw	at least 1	at least 2	at least 3	at least 4	at least 5	at least 6
LB	100%	0.07	0.02	0.01	0	0	0
	200%	0	0	0	0	0	0
MEAN	100%	0.35	0.14	0.09	0.06	0.04	0.03
	200%	0.07	0.03	0.01	0.01	0	0
UB	100%	1.41	0.81	0.58	0.45	0.37	0.31
	200%	0.65	0.38	0.28	0.22	0.19	0.16

Table J-7. APEX estimates for percent of children and adults with asthma in Tulsa study area, 2011.

Percent of children with asthma at elevated ventilation having exposures at or above 5-minute benchmark concentrations						
	number of days per year					
benchmark	at least 1	at least 2	at least 3	at least 4	at least 5	at least 6
100 ppb	0.24	0	0	0	0	0
200 ppb	0	0	0	0	0	0
300 ppb	0	0	0	0	0	0
400 ppb	0	0	0	0	0	0

Percent of children with asthma estimated to experience at least one day with an increase in sRaw ≥ 100%							
		number of days per year					
E-R Function	sRaw	at least 1	at least 2	at least 3	at least 4	at least 5	at least 6
LB	100%	0	0	0	0	0	0
	200%	0	0	0	0	0	0
MEAN	100%	0	0	0	0	0	0
	200%	0	0	0	0	0	0
UB	100%	0.44	0.24	0.20	0.16	0.11	0.11
	200%	0.20	0.11	0.09	0.07	0.05	0.05

Percent of adults with asthma at elevated ventilation having exposures at or above 5-minute benchmark concentrations						
	number of days per year					
benchmark	at least 1	at least 2	at least 3	at least 4	at least 5	at least 6
100 ppb	0.10	0	0	0	0	0
200 ppb	0	0	0	0	0	0
300 ppb	0	0	0	0	0	0
400 ppb	0	0	0	0	0	0

Percent of adults with asthma estimated to experience at least one day with an increase in sRaw ≥ 100%							
		number of days per year					
E-R Function	sRaw	at least 1	at least 2	at least 3	at least 4	at least 5	at least 6
LB	100%	0	0	0	0	0	0
	200%	0	0	0	0	0	0
MEAN	100%	0.01	0	0	0	0	0
	200%	0	0	0	0	0	0
UB	100%	0.21	0.09	0.06	0.03	0.03	0.01
	200%	0.10	0.05	0.03	0.01	0.01	0.01

Table J-8. APEX estimates for percent of children and adults with asthma in Tulsa study area, 2012.

Percent of children with asthma at elevated ventilation having exposures at or above 5-minute benchmark concentrations							
		number of days per year					
benchmark		at least 1	at least 2	at least 3	at least 4	at least 5	at least 6
100 ppb		0.15	0	0	0	0	0
200 ppb		0	0	0	0	0	0
300 ppb		0	0	0	0	0	0
400 ppb		0	0	0	0	0	0
Percent of children with asthma estimated to experience at least one day with an increase in sRaw ≥ 100%							
		number of days per year					
E-R Function	sRaw	at least 1	at least 2	at least 3	at least 4	at least 5	at least 6
LB	100%	0	0	0	0	0	0
	200%	0	0	0	0	0	0
MEAN	100%	0.02	0	0	0	0	0
	200%	0	0	0	0	0	0
UB	100%	0.53	0.35	0.26	0.22	0.18	0.16
	200%	0.22	0.18	0.13	0.11	0.09	0.07
Percent of adults with asthma at elevated ventilation having exposures at or above 5-minute benchmark concentrations							
		number of days per year					
benchmark		at least 1	at least 2	at least 3	at least 4	at least 5	at least 6
100 ppb		0.06	0	0	0	0	0
200 ppb		0	0	0	0	0	0
300 ppb		0	0	0	0	0	0
400 ppb		0	0	0	0	0	0
Percent of adults with asthma estimated to experience at least one day with an increase in sRaw ≥ 100%							
		number of days per year					
E-R Function	sRaw	at least 1	at least 2	at least 3	at least 4	at least 5	at least 6
LB	100%	0	0	0	0	0	0
	200%	0	0	0	0	0	0
MEAN	100%	0.01	0	0	0	0	0
	200%	0	0	0	0	0	0
UB	100%	0.23	0.11	0.07	0.05	0.03	0.02
	200%	0.11	0.06	0.03	0.02	0.02	0.01

Table J-9. APEX estimates for percent of children and adults with asthma in Tulsa study area, 2013.

Percent of children with asthma at elevated ventilation having exposures at or above 5-minute benchmark concentrations							
		number of days per year					
benchmark		at least 1	at least 2	at least 3	at least 4	at least 5	at least 6
100 ppb		0.03	0	0	0	0	0
200 ppb		0	0	0	0	0	0
300 ppb		0	0	0	0	0	0
400 ppb		0	0	0	0	0	0
Percent of children with asthma estimated to experience at least one day with an increase in sRaw ≥ 100%							
		number of days per year					
E-R Function	sRaw	at least 1	at least 2	at least 3	at least 4	at least 5	at least 6
LB	100%	0	0	0	0	0	0
	200%	0	0	0	0	0	0
MEAN	100%	0.02	0	0	0	0	0
	200%	0	0	0	0	0	0
UB	100%	0.44	0.33	0.27	0.22	0.18	0.16
	200%	0.20	0.15	0.15	0.09	0.09	0.07
Percent of adults with asthma at elevated ventilation having exposures at or above 5-minute benchmark concentrations							
		number of days per year					
benchmark		at least 1	at least 2	at least 3	at least 4	at least 5	at least 6
100 ppb		0	0	0	0	0	0
200 ppb		0	0	0	0	0	0
300 ppb		0	0	0	0	0	0
400 ppb		0	0	0	0	0	0
Percent of adults with asthma estimated to experience at least one day with an increase in sRaw ≥ 100%							
		number of days per year					
E-R Function	sRaw	at least 1	at least 2	at least 3	at least 4	at least 5	at least 6
LB	100%	0	0	0	0	0	0
	200%	0	0	0	0	0	0
MEAN	100%	0.01	0	0	0	0	0
	200%	0	0	0	0	0	0
UB	100%	0.19	0.09	0.05	0.04	0.03	0.02
	200%	0.09	0.05	0.03	0.02	0.01	0.01

Table J-10. Estimated daily maximum SO_2 exposures for air quality adjusted to just meet existing standard, while at elevated ventilation (binned): Fall River, 2011, children.

POPULATION ADJUSTED EXPOSURE TO BINS (NUMBER OF PEOPLE)						
Level	At least 1 Exposure	At least 2 Exposures	At least 3 Exposures	At least 4 Exposures	At least 5 Exposures	At least 6 Exposures
0	44	84	131	159	208	269
10	103	189	255	334	383	437
20	149	233	309	387	477	491
30	143	269	360	438	482	554
40	190	298	422	481	513	559
50	249	436	465	516	549	521
60	345	428	510	503	450	364
70	346	447	427	346	253	219
80	477	463	337	233	182	121
90	396	334	206	129	72	56
100	379	204	118	57	34	21
110	271	106	42	22	13	1
120	196	65	29	12	0	1
130	149	39	6	1	1	0
140	70	14	1	0	0	0
150	75	11	2	1	0	0
170	36	4	1	0	0	0
190	8	0	0	0	0	0
200	5	0	0	0	0	0
210	3	0	0	0	0	0
230	0	0	0	0	0	0
250	0	0	0	0	0	0
300	0	0	0	0	0	0
350	0	0	0	0	0	0
400	0	0	0	0	0	0
450	0	0	0	0	0	0
500	0	0	0	0	0	0
550	0	0	0	0	0	0
600	0	0	0	0	0	0

Table J-11. Estimated daily maximum SO_2 exposures for air quality adjusted to just meet existing standard, while at elevated ventilation (binned): Fall River, 2011, adults.

POPULATION ADJUSTED EXPOSURE TO BINS (NUMBER OF PEOPLE)						
Level	At least 1 Exposure	At least 2 Exposures	At least 3 Exposures	At least 4 Exposures	At least 5 Exposures	At least 6 Exposures
0	846	1960	3040	4178	5175	6026
10	2399	4026	4673	4712	4524	4225
20	2302	2473	2192	1828	1445	1173
30	1690	1417	1080	768	586	435
40	1257	900	550	376	251	167
50	995	604	333	162	115	104
60	740	327	184	97	76	39
70	554	238	82	69	17	11
80	521	167	52	19	9	2
90	327	71	32	4	2	2
100	236	30	6	0	0	0
110	154	15	0	0	0	0
120	87	9	0	0	0	0
130	63	0	0	0	0	0
140	37	0	0	0	0	0
150	22	0	0	0	0	0
170	24	0	0	0	0	0
190	2	0	0	0	0	0
200	2	0	0	0	0	0
210	0	0	0	0	0	0
230	0	0	0	0	0	0
250	0	0	0	0	0	0
300	0	0	0	0	0	0
350	0	0	0	0	0	0
400	0	0	0	0	0	0
450	0	0	0	0	0	0
500	0	0	0	0	0	0
550	0	0	0	0	0	0
600	0	0	0	0	0	0

Table J-12. Estimated daily maximum SO_2 exposures for air quality adjusted to just meet existing standard, while at elevated ventilation (binned): Fall River, 2012, children.

POPULATION ADJUSTED EXPOSURE TO BINS (NUMBER OF PEOPLE)						
Level	At least 1 Exposure	At least 2 Exposures	At least 3 Exposures	At least 4 Exposures	At least 5 Exposures	At least 6 Exposures
0	56	107	163	213	273	334
10	120	252	338	428	490	543
20	183	310	420	510	564	630
30	266	411	552	610	738	828
40	350	518	636	724	694	651
50	375	546	539	479	423	350
60	522	551	495	386	296	191
70	513	465	281	191	98	67
80	400	219	122	53	34	21
90	366	150	58	26	11	5
100	391	93	19	4	1	0
110	66	5	1	0	0	0
120	13	1	0	0	0	0
130	5	1	0	0	0	0
140	3	0	0	0	0	0
150	2	0	0	0	0	0
170	0	0	0	0	0	0
190	0	0	0	0	0	0
200	0	0	0	0	0	0
210	0	0	0	0	0	0
230	0	0	0	0	0	0
250	0	0	0	0	0	0
300	0	0	0	0	0	0
350	0	0	0	0	0	0
400	0	0	0	0	0	0
450	0	0	0	0	0	0
500	0	0	0	0	0	0
550	0	0	0	0	0	0
600	0	0	0	0	0	0

Table J-13. Estimated daily maximum SO_2 exposures for air quality adjusted to just meet existing standard, while at elevated ventilation (binned): Fall River, 2012, adults.

POPULATION ADJUSTED EXPOSURE TO BINS (NUMBER OF PEOPLE)						
Level	At least 1 Exposure	At least 2 Exposures	At least 3 Exposures	At least 4 Exposures	At least 5 Exposures	At least 6 Exposures
0	1181	2616	4018	5093	6194	7068
10	3038	4422	4656	4604	4189	3734
20	2562	2475	1954	1523	1186	956
30	1770	1287	883	591	398	275
40	1225	666	379	249	143	97
50	764	353	203	102	65	32
60	608	216	91	35	17	17
70	411	123	32	22	13	4
80	273	45	15	4	0	0
90	199	17	4	0	0	0
100	214	22	0	0	0	0
110	11	0	0	0	0	0
120	2	0	0	0	0	0
130	2	0	0	0	0	0
140	0	0	0	0	0	0
150	0	0	0	0	0	0
170	0	0	0	0	0	0
190	0	0	0	0	0	0
200	0	0	0	0	0	0
210	0	0	0	0	0	0
230	0	0	0	0	0	0
250	0	0	0	0	0	0
300	0	0	0	0	0	0
350	0	0	0	0	0	0
400	0	0	0	0	0	0
450	0	0	0	0	0	0
500	0	0	0	0	0	0
550	0	0	0	0	0	0
600	0	0	0	0	0	0

Table J-14. Estimated daily maximum SO_2 exposures for air quality adjusted to just meet existing standard, while at elevated ventilation (binned): Fall River, 2013, children.

POPULATION ADJUSTED EXPOSURE TO BINS (NUMBER OF PEOPLE)						
Level	At least 1 Exposure	At least 2 Exposures	At least 3 Exposures	At least 4 Exposures	At least 5 Exposures	At least 6 Exposures
0	38	101	132	176	239	294
10	173	273	403	494	592	671
20	419	644	819	1006	1089	1180
30	453	718	851	885	917	895
40	706	884	878	763	608	466
50	560	513	333	199	132	89
60	365	245	131	67	27	16
70	166	93	38	19	8	3
80	180	63	18	6	3	1
90	125	36	9	3	1	0
100	193	41	11	1	0	0
110	97	14	1	0	0	0
120	149	2	0	0	0	0
130	4	1	0	0	0	0
140	2	0	0	0	0	0
150	1	0	0	0	0	0
170	0	0	0	0	0	0
190	0	0	0	0	0	0
200	0	0	0	0	0	0
210	0	0	0	0	0	0
230	0	0	0	0	0	0
250	0	0	0	0	0	0
300	0	0	0	0	0	0
350	0	0	0	0	0	0
400	0	0	0	0	0	0
450	0	0	0	0	0	0
500	0	0	0	0	0	0
550	0	0	0	0	0	0
600	0	0	0	0	0	0

Table J-15. Estimated daily maximum SO_2 exposures for air quality adjusted to just meet existing standard, while at elevated ventilation (binned): Fall River, 2013, adults.

POPULATION ADJUSTED EXPOSURE TO BINS (NUMBER OF PEOPLE)						
Level	At least 1 Exposure	At least 2 Exposures	At least 3 Exposures	At least 4 Exposures	At least 5 Exposures	At least 6 Exposures
0	1190	2540	3819	4922	5889	6649
10	3914	5240	5344	5184	4788	4422
20	3375	2914	2296	1685	1246	943
30	1597	948	508	314	206	134
40	1077	385	195	91	61	35
50	534	117	50	15	11	4
60	208	52	11	9	0	0
70	97	13	2	0	0	0
80	56	17	2	0	0	0
90	43	4	2	0	0	0
100	74	9	0	0	0	0
110	50	0	0	0	0	0
120	35	0	0	0	0	0
130	4	0	0	0	0	0
140	0	0	0	0	0	0
150	0	0	0	0	0	0
170	0	0	0	0	0	0
190	0	0	0	0	0	0
200	0	0	0	0	0	0
210	0	0	0	0	0	0
230	0	0	0	0	0	0
250	0	0	0	0	0	0
300	0	0	0	0	0	0
350	0	0	0	0	0	0
400	0	0	0	0	0	0
450	0	0	0	0	0	0
500	0	0	0	0	0	0
550	0	0	0	0	0	0
600	0	0	0	0	0	0

Table J-16. Estimated daily maximum SO_2 exposures for air quality adjusted to just meet existing standard, while at elevated ventilation (binned): Indianapolis, 2011, children.

POPULATION ADJUSTED EXPOSURE TO BINS (NUMBER OF PEOPLE)						
Level	At least 1 Exposure	At least 2 Exposures	At least 3 Exposures	At least 4 Exposures	At least 5 Exposures	At least 6 Exposures
0	34	56	64	86	105	142
10	116	213	378	502	685	802
20	292	577	764	914	955	1082
30	401	678	794	985	1154	1240
40	667	1067	1446	1700	1985	2202
50	1015	1566	2030	2375	2487	2573
60	1258	1899	2015	1959	1850	1693
70	1345	1584	1472	1157	925	693
80	1708	1517	1105	794	483	292
90	1064	802	476	228	124	60
100	652	315	127	86	37	7
110	1337	434	124	7	7	4
120	558	79	15	11	0	0
130	142	19	0	0	0	0
140	60	11	0	0	0	0
150	7	0	0	0	0	0
170	64	7	0	0	0	0
190	0	0	0	0	0	0
200	0	0	0	0	0	0
210	22	0	0	0	0	0
230	0	0	0	0	0	0
250	0	0	0	0	0	0
300	0	0	0	0	0	0
350	56	0	0	0	0	0
400	34	0	0	0	0	0
450	0	0	0	0	0	0
500	0	0	0	0	0	0
550	0	0	0	0	0	0
600	0	0	0	0	0	0

Table J-17. Estimated daily maximum SO_2 exposures for air quality adjusted to just meet existing standard, while at elevated ventilation (binned): Indianapolis, 2011, adults.

POPULATION ADJUSTED EXPOSURE TO BINS (NUMBER OF PEOPLE)						
Level	At least 1 Exposure	At least 2 Exposures	At least 3 Exposures	At least 4 Exposures	At least 5 Exposures	At least 6 Exposures
0	398	1014	1749	2701	3796	4860
10	3211	6683	9807	12203	13952	15240
20	5601	8600	9708	9764	9477	8911
30	6198	6939	6578	5675	4705	3802
40	6422	6067	4381	3479	2570	2184
50	4885	3267	2222	1431	996	647
60	3211	1624	902	454	268	149
70	1917	921	336	143	106	68
80	1767	573	218	100	44	37
90	915	143	68	25	6	0
100	467	87	25	0	0	0
110	697	100	12	0	0	0
120	149	12	0	0	0	0
130	100	19	0	0	0	0
140	37	0	0	0	0	0
150	25	0	0	0	0	0
170	25	0	0	0	0	0
190	6	0	0	0	0	0
200	0	0	0	0	0	0
210	12	0	0	0	0	0
230	0	0	0	0	0	0
250	0	0	0	0	0	0
300	0	0	0	0	0	0
350	6	0	0	0	0	0
400	25	0	0	0	0	0
450	0	0	0	0	0	0
500	0	0	0	0	0	0
550	0	0	0	0	0	0
600	0	0	0	0	0	0

Table J-18. Estimated daily maximum SO_2 exposures for air quality adjusted to just meet existing standard, while at elevated ventilation (binned): Indianapolis, 2012, children.

POPULATION ADJUSTED EXPOSURE TO BINS (NUMBER OF PEOPLE)						
Level	At least 1 Exposure	At least 2 Exposures	At least 3 Exposures	At least 4 Exposures	At least 5 Exposures	At least 6 Exposures
0	15	56	67	112	124	135
10	127	292	449	622	779	910
20	296	509	757	888	981	1116
30	378	659	832	929	1127	1184
40	607	963	1150	1408	1543	1828
50	1101	1637	2109	2416	2659	2644
60	1626	2274	2453	2375	2210	1963
70	1723	1835	1558	1281	910	719
80	1277	1075	764	442	273	213
90	1258	697	408	247	161	71
100	779	461	165	67	37	19
110	461	202	82	30	4	0
120	607	127	22	4	4	4
130	127	7	4	4	0	0
140	109	19	4	0	0	0
150	337	15	0	0	0	0
170	0	0	0	0	0	0
190	0	0	0	0	0	0
200	0	0	0	0	0	0
210	0	0	0	0	0	0
230	0	0	0	0	0	0
250	0	0	0	0	0	0
300	0	0	0	0	0	0
350	0	0	0	0	0	0
400	0	0	0	0	0	0
450	0	0	0	0	0	0
500	0	0	0	0	0	0
550	0	0	0	0	0	0
600	0	0	0	0	0	0

Table J-19. Estimated daily maximum SO_2 exposures for air quality adjusted to just meet existing standard, while at elevated ventilation (binned): Indianapolis, 2012, adults.

POPULATION ADJUSTED EXPOSURE TO BINS (NUMBER OF PEOPLE)						
Level	At least 1 Exposure	At least 2 Exposures	At least 3 Exposures	At least 4 Exposures	At least 5 Exposures	At least 6 Exposures
0	429	896	1593	2651	3672	4736
10	3460	7449	10784	13180	14991	16292
20	5881	8202	9154	9135	8805	8084
30	5489	6690	6030	5308	4524	3933
40	5986	5420	4232	3223	2365	1755
50	5196	3790	2732	1649	1108	803
60	4207	2122	915	516	361	199
70	2197	765	292	180	68	50
80	1058	311	137	93	56	37
90	821	218	100	44	6	6
100	554	118	31	6	6	6
110	380	44	12	12	6	0
120	236	37	6	0	0	0
130	56	0	0	0	0	0
140	50	0	0	0	0	0
150	93	0	0	0	0	0
170	0	0	0	0	0	0
190	0	0	0	0	0	0
200	0	0	0	0	0	0
210	0	0	0	0	0	0
230	0	0	0	0	0	0
250	0	0	0	0	0	0
300	0	0	0	0	0	0
350	0	0	0	0	0	0
400	0	0	0	0	0	0
450	0	0	0	0	0	0
500	0	0	0	0	0	0
550	0	0	0	0	0	0
600	0	0	0	0	0	0

Table J-20. Estimated daily maximum SO_2 exposures for air quality adjusted to just meet existing standard, while at elevated ventilation (binned): Indianapolis, 2013, children.

POPULATION ADJUSTED EXPOSURE TO BINS (NUMBER OF PEOPLE)						
Level	At least 1 Exposure	At least 2 Exposures	At least 3 Exposures	At least 4 Exposures	At least 5 Exposures	At least 6 Exposures
0	30	79	94	127	139	161
10	146	228	401	581	742	903
20	243	536	738	895	981	1127
30	472	757	963	1015	1176	1169
40	712	1240	1412	1802	2041	2363
50	1401	2060	2614	2959	3083	2993
60	1749	2468	2382	1989	1704	1476
70	1940	1513	1247	925	689	461
80	1300	933	524	326	154	90
90	884	498	255	150	71	37
100	880	285	135	26	7	0
110	408	146	37	4	7	7
120	131	49	15	4	0	0
130	64	4	0	0	0	0
140	311	26	0	0	0	0
150	45	0	0	0	0	0
170	0	0	0	0	0	0
190	7	0	0	0	0	0
200	52	0	0	0	0	0
210	0	0	0	0	0	0
230	49	0	0	0	0	0
250	0	0	0	0	0	0
300	0	0	0	0	0	0
350	0	0	0	0	0	0
400	0	0	0	0	0	0
450	0	0	0	0	0	0
500	0	0	0	0	0	0
550	0	0	0	0	0	0
600	0	0	0	0	0	0

Table J-21. Estimated daily maximum SO_2 exposures for air quality adjusted to just meet existing standard, while at elevated ventilation (binned): Indianapolis, 2013, adults.

POPULATION ADJUSTED EXPOSURE TO BINS (NUMBER OF PEOPLE)						
Level	At least 1 Exposure	At least 2 Exposures	At least 3 Exposures	At least 4 Exposures	At least 5 Exposures	At least 6 Exposures
0	454	1164	2159	3379	4499	5569
10	3522	7648	10884	13168	14661	15899
20	5924	8301	8830	8662	8507	7835
30	6254	7331	6590	5557	4536	3952
40	6932	5688	4269	3279	2464	1886
50	5345	3342	2116	1388	940	523
60	3105	1369	728	324	187	143
70	1774	616	249	131	87	50
80	1151	317	87	44	37	25
90	554	112	37	19	6	0
100	454	87	25	6	0	0
110	243	19	6	0	0	0
120	106	12	6	6	0	0
130	19	0	0	0	0	0
140	93	6	6	0	0	0
150	68	6	0	0	0	0
170	0	0	0	0	0	0
190	6	0	0	0	0	0
200	25	0	0	0	0	0
210	19	0	0	0	0	0
230	19	0	0	0	0	0
250	0	0	0	0	0	0
300	0	0	0	0	0	0
350	0	0	0	0	0	0
400	0	0	0	0	0	0
450	0	0	0	0	0	0
500	0	0	0	0	0	0
550	0	0	0	0	0	0
600	0	0	0	0	0	0

Table J-22. Estimated daily maximum SO_2 exposures for air quality adjusted to just meet existing standard, while at elevated ventilation (binned): Tulsa, 2011, children.

POPULATION ADJUSTED EXPOSURE TO BINS (NUMBER OF PEOPLE)						
Level	At least 1 Exposure	At least 2 Exposures	At least 3 Exposures	At least 4 Exposures	At least 5 Exposures	At least 6 Exposures
0	224	460	724	930	1168	1397
10	1679	2616	3040	3251	3281	3218
20	1887	1570	1166	881	698	589
30	807	452	302	266	218	181
40	429	228	167	99	66	49
50	223	104	49	23	21	13
60	119	23	8	8	5	5
70	48	8	5	2	0	0
80	20	2	0	0	0	0
90	16	0	0	0	0	0
100	8	0	0	0	0	0
110	2	0	0	0	0	0
120	2	0	0	0	0	0
130	0	0	0	0	0	0
140	2	0	0	0	0	0
150	0	0	0	0	0	0
170	0	0	0	0	0	0
190	0	0	0	0	0	0
200	0	0	0	0	0	0
210	0	0	0	0	0	0
230	0	0	0	0	0	0
250	0	0	0	0	0	0
300	0	0	0	0	0	0
350	0	0	0	0	0	0
400	0	0	0	0	0	0
450	0	0	0	0	0	0
500	0	0	0	0	0	0
550	0	0	0	0	0	0
600	0	0	0	0	0	0

Table J-23. Estimated daily maximum SO_2 exposures for air quality adjusted to just meet existing standard, while at elevated ventilation (binned): Tulsa, 2011, adults.

POPULATION ADJUSTED EXPOSURE TO BINS (NUMBER OF PEOPLE)						
Level	At least 1 Exposure	At least 2 Exposures	At least 3 Exposures	At least 4 Exposures	At least 5 Exposures	At least 6 Exposures
0	4898	7860	9613	10849	11728	12248
10	6176	5478	4411	3487	2783	2341
20	2272	1052	618	437	306	244
30	772	333	214	134	89	59
40	437	163	86	36	18	15
50	258	74	24	9	6	0
60	92	15	0	0	0	0
70	50	3	0	0	0	0
80	9	0	0	0	0	0
90	12	0	0	0	0	0
100	6	0	0	0	0	0
110	6	0	0	0	0	0
120	0	0	0	0	0	0
130	3	0	0	0	0	0
140	0	0	0	0	0	0
150	0	0	0	0	0	0
170	0	0	0	0	0	0
190	0	0	0	0	0	0
200	0	0	0	0	0	0
210	0	0	0	0	0	0
230	0	0	0	0	0	0
250	0	0	0	0	0	0
300	0	0	0	0	0	0
350	0	0	0	0	0	0
400	0	0	0	0	0	0
450	0	0	0	0	0	0
500	0	0	0	0	0	0
550	0	0	0	0	0	0
600	0	0	0	0	0	0

Table J-24. Estimated daily maximum SO_2 exposures for air quality adjusted to just meet existing standard, while at elevated ventilation (binned): Tulsa, 2012, children.

POPULATION ADJUSTED EXPOSURE TO BINS (NUMBER OF PEOPLE)						
Level	At least 1 Exposure	At least 2 Exposures	At least 3 Exposures	At least 4 Exposures	At least 5 Exposures	At least 6 Exposures
0	203	437	670	882	1105	1285
10	967	1547	1940	2209	2352	2444
20	2397	2476	2133	1823	1555	1353
30	965	551	432	351	307	284
40	607	356	238	175	130	81
50	147	76	46	18	7	5
60	92	15	2	0	0	0
70	38	3	0	0	0	0
80	30	0	0	0	0	0
90	15	0	0	0	0	0
100	2	0	0	0	0	0
110	0	0	0	0	0	0
120	2	0	0	0	0	0
130	0	0	0	0	0	0
140	2	0	0	0	0	0
150	3	0	0	0	0	0
170	0	0	0	0	0	0
190	0	0	0	0	0	0
200	0	0	0	0	0	0
210	0	0	0	0	0	0
230	0	0	0	0	0	0
250	0	0	0	0	0	0
300	0	0	0	0	0	0
350	0	0	0	0	0	0
400	0	0	0	0	0	0
450	0	0	0	0	0	0
500	0	0	0	0	0	0
550	0	0	0	0	0	0
600	0	0	0	0	0	0

Table J-25. Estimated daily maximum SO_2 exposures for air quality adjusted to just meet existing standard, while at elevated ventilation (binned): Tulsa, 2012, adults.

POPULATION ADJUSTED EXPOSURE TO BINS (NUMBER OF PEOPLE)						
Level	At least 1 Exposure	At least 2 Exposures	At least 3 Exposures	At least 4 Exposures	At least 5 Exposures	At least 6 Exposures
0	4474	7260	8998	10201	11057	11633
10	5228	5017	4331	3615	3027	2611
20	3571	2020	1251	867	659	496
30	957	413	258	172	119	107
40	496	214	92	53	33	21
50	140	30	9	6	0	0
60	65	9	0	0	0	0
70	21	0	0	0	0	0
80	18	0	0	0	0	0
90	6	0	0	0	0	0
100	0	0	0	0	0	0
110	0	0	0	0	0	0
120	0	0	0	0	0	0
130	3	0	0	0	0	0
140	3	0	0	0	0	0
150	0	0	0	0	0	0
170	3	0	0	0	0	0
190	0	0	0	0	0	0
200	0	0	0	0	0	0
210	0	0	0	0	0	0
230	0	0	0	0	0	0
250	0	0	0	0	0	0
300	0	0	0	0	0	0
350	0	0	0	0	0	0
400	0	0	0	0	0	0
450	0	0	0	0	0	0
500	0	0	0	0	0	0
550	0	0	0	0	0	0
600	0	0	0	0	0	0

Table J-26. Estimated daily maximum SO_2 exposures for air quality adjusted to just meet existing standard, while at elevated ventilation (binned): Tulsa, 2013, children.

POPULATION ADJUSTED EXPOSURE TO BINS (NUMBER OF PEOPLE)						
Level	At least 1 Exposure	At least 2 Exposures	At least 3 Exposures	At least 4 Exposures	At least 5 Exposures	At least 6 Exposures
0	300	582	813	1034	1234	1417
10	815	1183	1465	1646	1823	1951
20	3009	2987	2710	2436	2146	1885
30	691	422	279	233	178	142
40	576	266	183	104	73	53
50	61	25	10	5	3	3
60	13	2	0	0	0	0
70	0	0	0	0	0	0
80	0	0	0	0	0	0
90	0	0	0	0	0	0
100	2	0	0	0	0	0
110	0	0	0	0	0	0
120	0	0	0	0	0	0
130	0	0	0	0	0	0
140	0	0	0	0	0	0
150	0	0	0	0	0	0
170	0	0	0	0	0	0
190	0	0	0	0	0	0
200	0	0	0	0	0	0
210	0	0	0	0	0	0
230	0	0	0	0	0	0
250	0	0	0	0	0	0
300	0	0	0	0	0	0
350	0	0	0	0	0	0
400	0	0	0	0	0	0
450	0	0	0	0	0	0
500	0	0	0	0	0	0
550	0	0	0	0	0	0
600	0	0	0	0	0	0

Table J-27. Estimated daily maximum SO_2 exposures for air quality adjusted to just meet existing standard, while at elevated ventilation (binned): Tulsa, 2013, adults.

POPULATION ADJUSTED EXPOSURE TO BINS (NUMBER OF PEOPLE)						
Level	At least 1 Exposure	At least 2 Exposures	At least 3 Exposures	At least 4 Exposures	At least 5 Exposures	At least 6 Exposures
0	5190	8264	9901	10997	11698	12251
10	4688	4108	3508	2956	2519	2127
20	4180	2326	1438	924	656	487
30	576	220	83	45	36	24
40	333	59	27	18	6	6
50	24	3	3	0	0	0
60	0	0	0	0	0	0
70	0	0	0	0	0	0
80	3	0	0	0	0	0
90	0	0	0	0	0	0
100	0	0	0	0	0	0
110	0	0	0	0	0	0
120	0	0	0	0	0	0
130	0	0	0	0	0	0
140	0	0	0	0	0	0
150	0	0	0	0	0	0
170	0	0	0	0	0	0
190	0	0	0	0	0	0
200	0	0	0	0	0	0
210	0	0	0	0	0	0
230	0	0	0	0	0	0
250	0	0	0	0	0	0
300	0	0	0	0	0	0
350	0	0	0	0	0	0
400	0	0	0	0	0	0
450	0	0	0	0	0	0
500	0	0	0	0	0	0
550	0	0	0	0	0	0
600	0	0	0	0	0	0

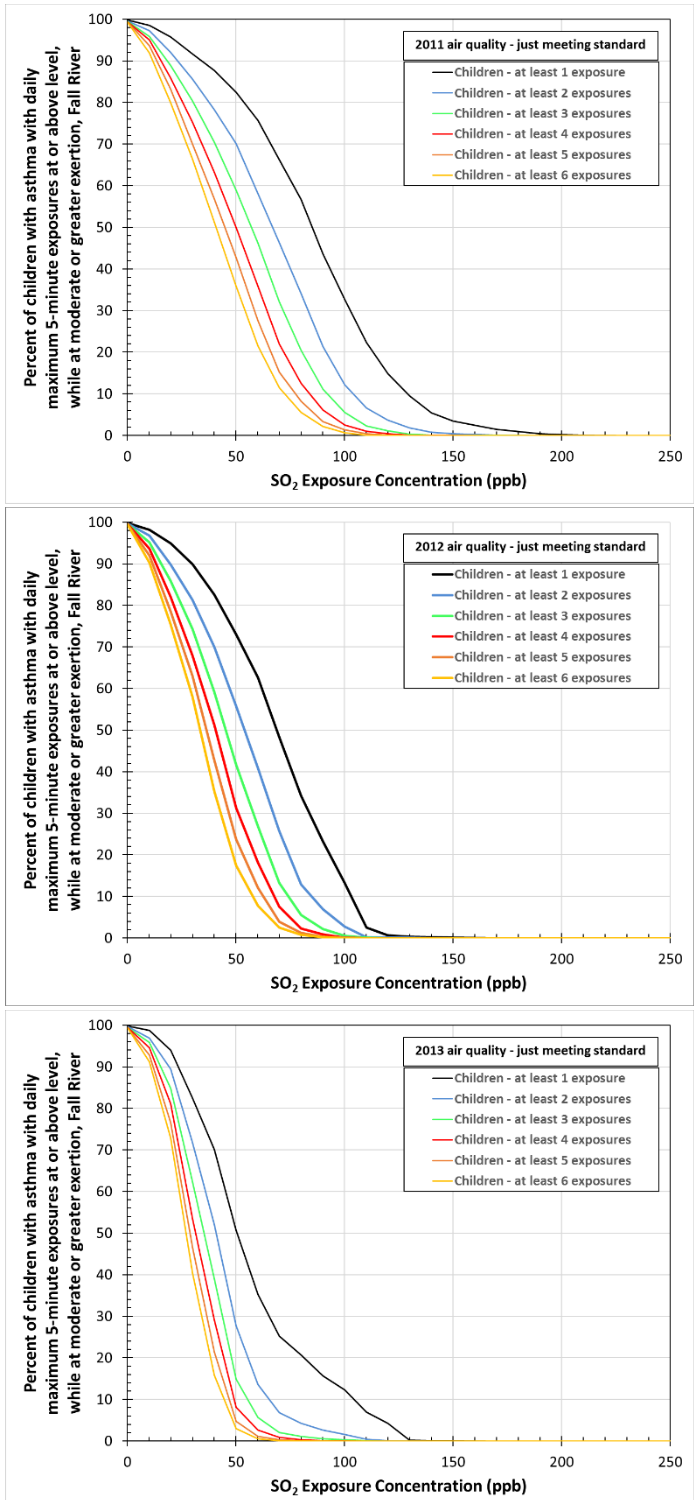


Figure J-1. Estimated percent of children with asthma expected to experience daily maximum 5-minute SO_2 exposures at or above selected levels in Fall River study area, air quality adjusted to just meet the existing standard, 2011-2013 (top to bottom panels).

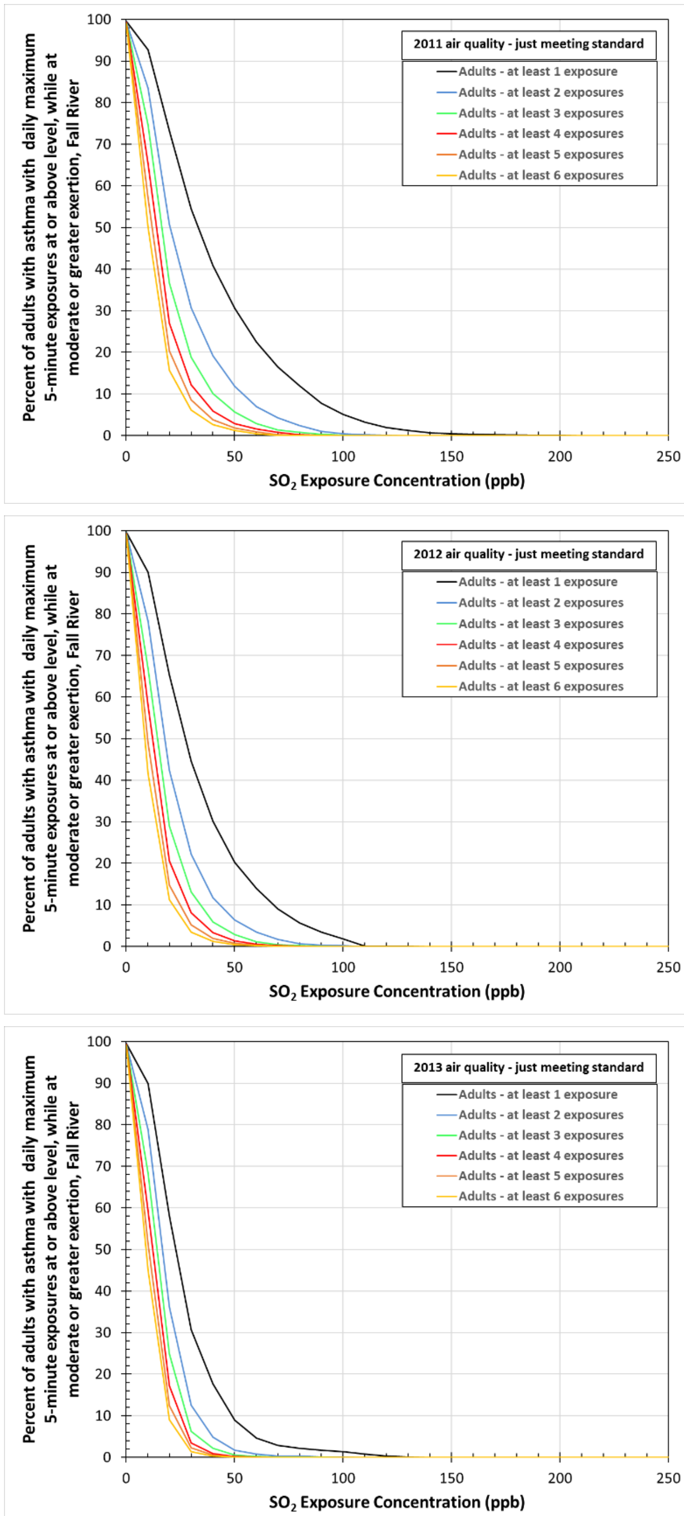


Figure J-2. Estimated percent of adults with asthma expected to experience daily maximum 5-minute SO_2 exposures at or above selected levels in Fall River study area, air quality adjusted to just meet the existing standard, 2011-2013 (top to bottom panels).

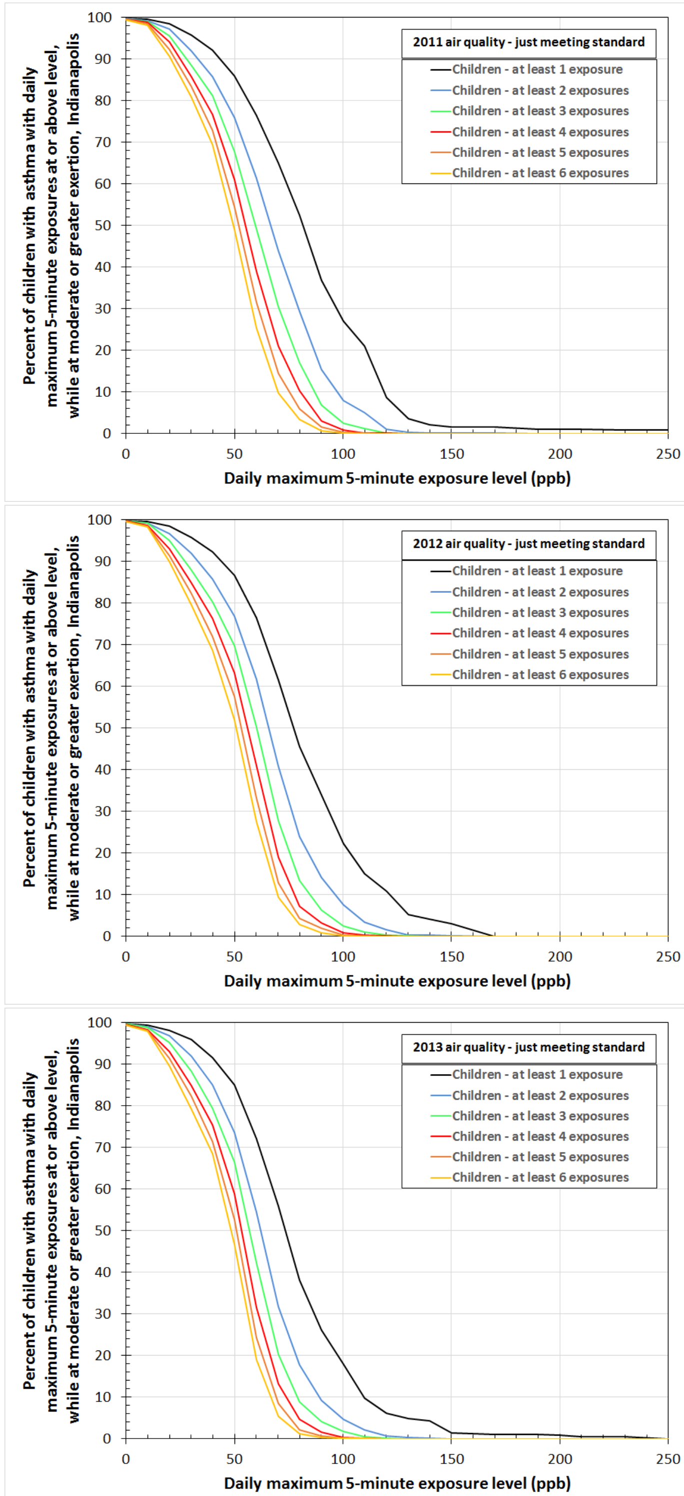


Figure J-3. Estimated percent of children with asthma expected to experience daily maximum 5-minute SO_2 exposures at or above selected levels in Indianapolis study area, air quality adjusted to just meet the existing standard, 2011-2013 (top to bottom panels).

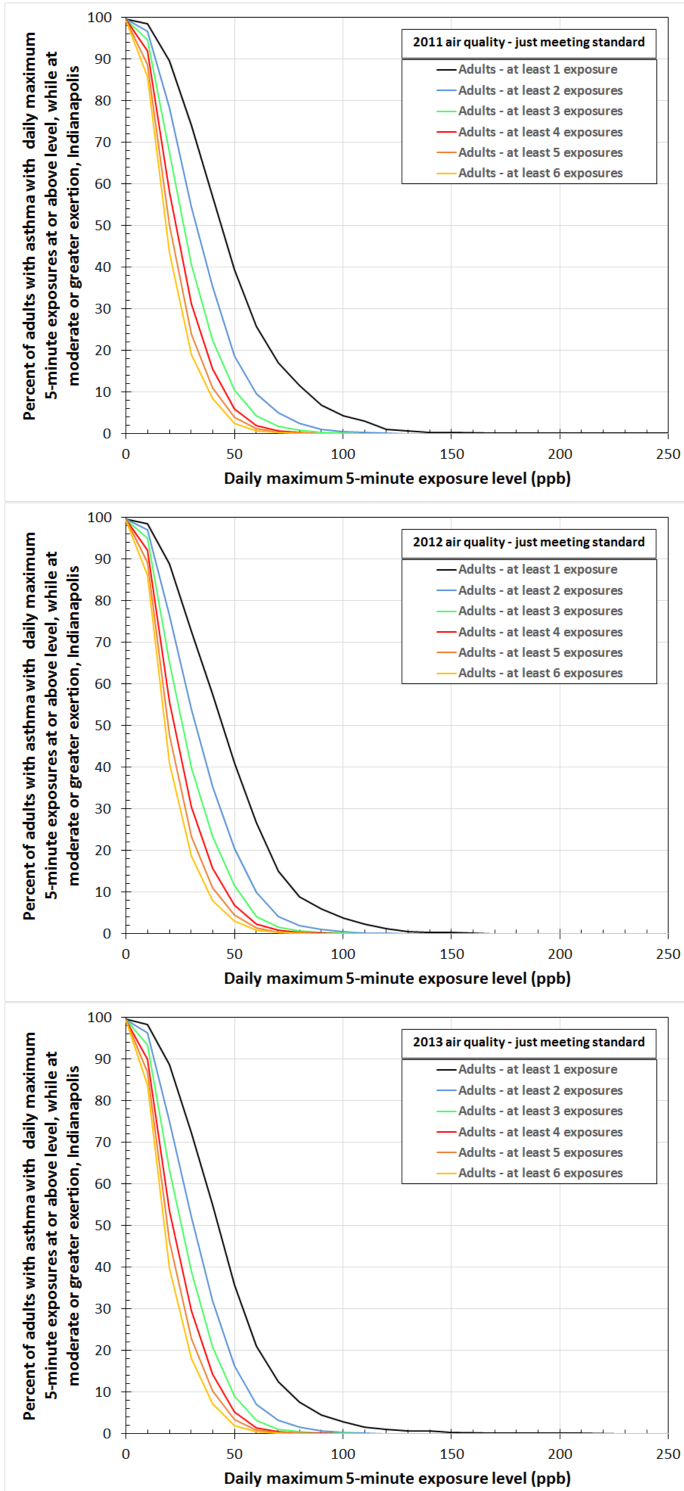


Figure J-4. Estimated percent of adults with asthma expected to experience daily maximum 5-minute SO_2 exposures at or above selected levels in Indianapolis study area, air quality adjusted to just meet the existing standard, 2011-2013 (top to bottom panels).

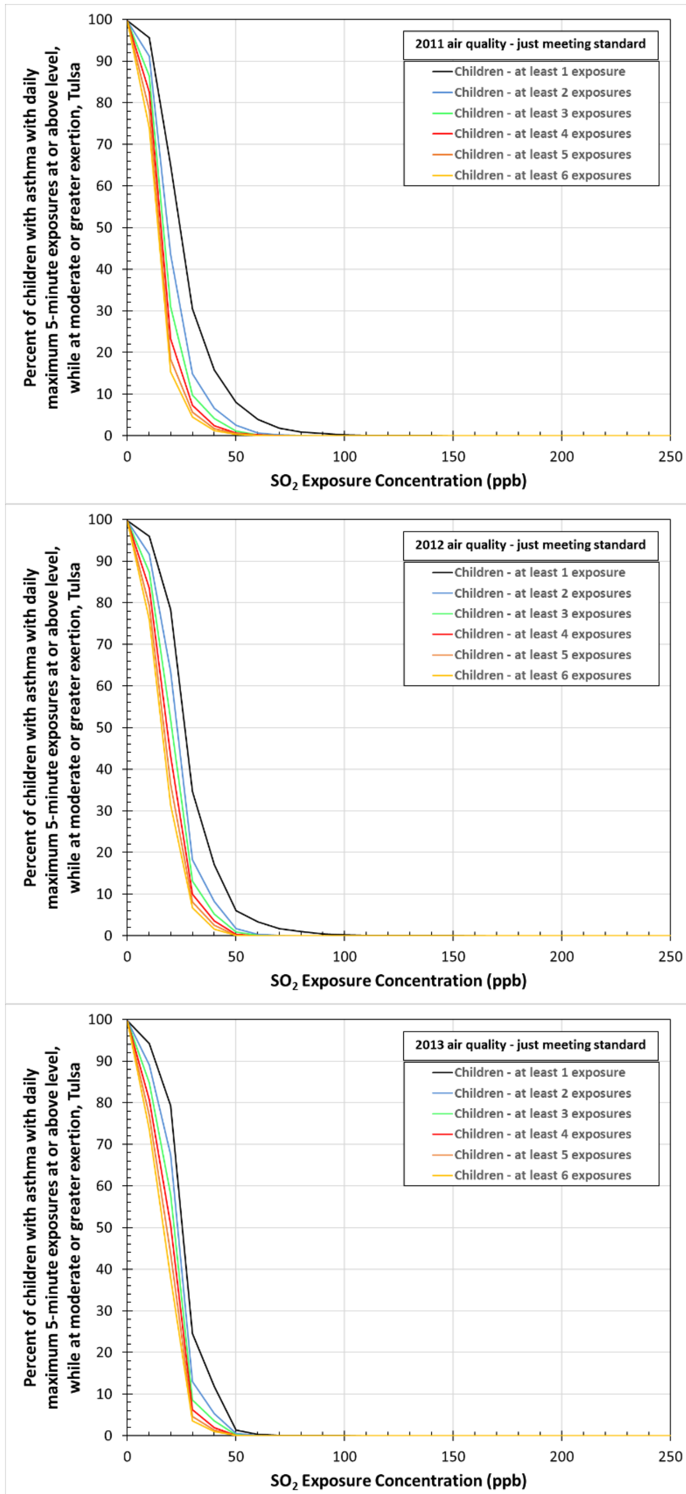


Figure J-5. Estimated percent of children with asthma expected to experience daily maximum 5-minute SO_2 exposures at or above selected levels in Tulsa study area, air quality adjusted to just meet the existing standard, 2011-2013 (top to bottom panels).

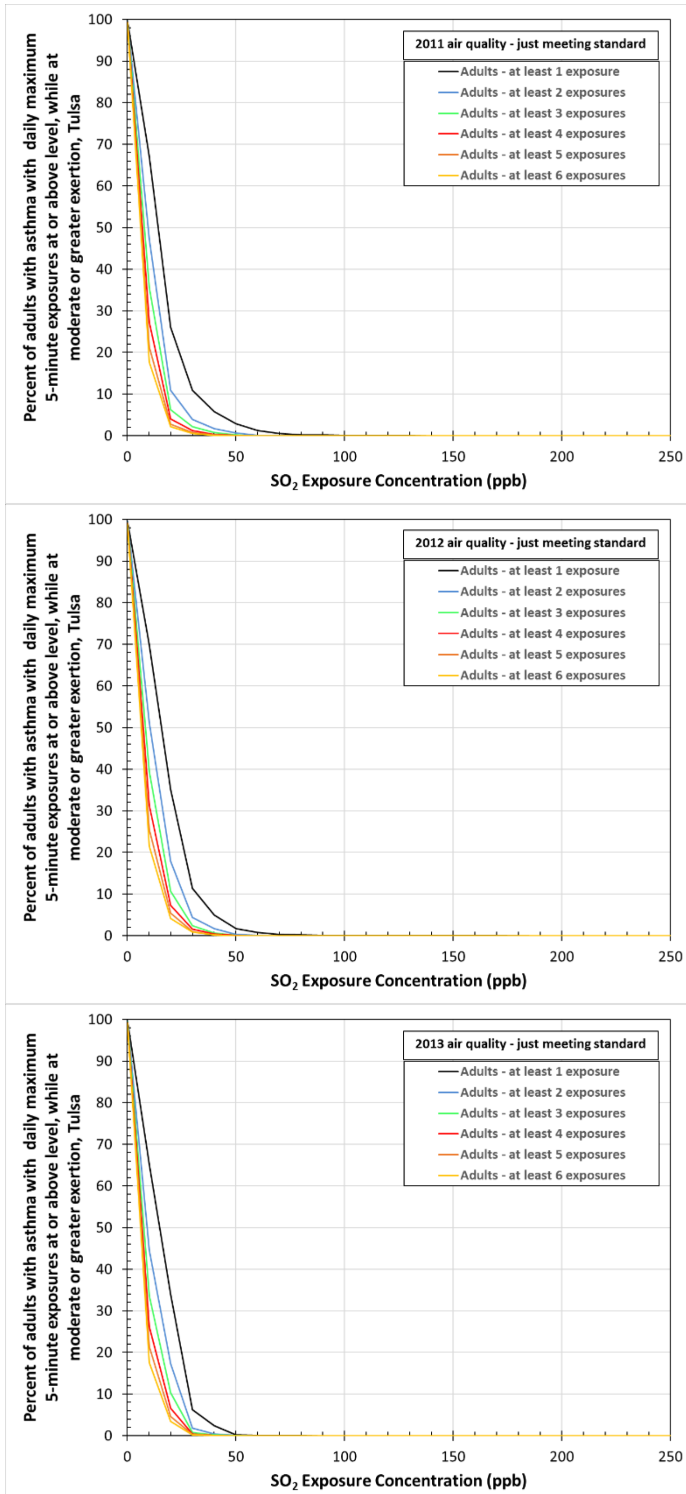


Figure J-6. Estimated percent of adults with asthma expected to experience daily maximum 5-minute SO_2 exposures at or above selected levels in Tulsa study area, air quality adjusted to just meet the existing standard, 2011-2013 (top to bottom panels).

Table J-28. Exposure-Response Function for SO2-attributable increases (≥100% and ≥200%) in sRaw: Mean, lower prediction interval and upper prediction interval.

E-R sRaw 100%				E-R sRaw 200%			
exposure	**mean**	**lower**	**upper**	**exposure**	**mean**	**lower**	**upper**
5	2.49E-07	2.87E-10	5.74E-05	**5**	5.77E-08	6.95E-12	6.09E-05
15	4.02E-05	6.70E-07	1.14E-03	**15**	8.64E-06	3.07E-08	7.35E-04
25	2.92E-04	1.33E-05	3.71E-03	**25**	6.38E-05	8.34E-07	2.04E-03
35	9.45E-04	7.74E-05	7.53E-03	**35**	2.13E-04	5.97E-06	3.81E-03
45	2.12E-03	2.58E-04	1.24E-02	**45**	4.93E-04	2.33E-05	5.93E-03
55	3.90E-03	6.33E-04	1.79E-02	**55**	9.32E-04	6.48E-05	8.32E-03
65	6.28E-03	1.28E-03	2.41E-02	**65**	1.55E-03	1.45E-04	1.09E-02
75	9.26E-03	2.25E-03	3.08E-02	**75**	2.34E-03	2.81E-04	1.37E-02
85	1.28E-02	3.61E-03	3.78E-02	**85**	3.33E-03	4.88E-04	1.66E-02
95	1.69E-02	5.40E-03	4.51E-02	**95**	4.50E-03	7.83E-04	1.96E-02
105	2.15E-02	7.64E-03	5.26E-02	**105**	5.86E-03	1.18E-03	2.27E-02
115	2.66E-02	1.03E-02	6.03E-02	**115**	7.40E-03	1.69E-03	2.59E-02
125	3.21E-02	1.35E-02	6.81E-02	**125**	9.11E-03	2.33E-03	2.92E-02
135	3.80E-02	1.71E-02	7.60E-02	**135**	1.10E-02	3.10E-03	3.25E-02
145	4.41E-02	2.12E-02	8.39E-02	**145**	1.30E-02	4.02E-03	3.58E-02
160	5.40E-02	2.81E-02	9.59E-02	**160**	1.63E-02	5.67E-03	4.09E-02
180	6.80E-02	3.87E-02	1.12E-01	**180**	2.13E-02	8.42E-03	4.79E-02
195	7.90E-02	4.76E-02	1.24E-01	**195**	2.53E-02	1.09E-02	5.31E-02
205	8.65E-02	5.39E-02	1.32E-01	**205**	2.81E-02	1.27E-02	5.66E-02
220	9.80E-02	6.38E-02	1.44E-01	**220**	3.25E-02	1.57E-02	6.19E-02
240	1.14E-01	7.79E-02	1.60E-01	**240**	3.87E-02	2.03E-02	6.91E-02
275	1.42E-01	1.04E-01	1.87E-01	**275**	5.04E-02	2.95E-02	8.17E-02
325	1.82E-01	1.44E-01	2.26E-01	**325**	6.83E-02	4.48E-02	1.00E-01
375	2.22E-01	1.83E-01	2.64E-01	**375**	8.72E-02	6.20E-02	1.20E-01
425	2.60E-01	2.20E-01	3.03E-01	**425**	1.07E-01	7.99E-02	1.40E-01
475	2.97E-01	2.55E-01	3.41E-01	**475**	1.27E-01	9.77E-02	1.61E-01
525	3.32E-01	2.87E-01	3.80E-01	**525**	1.47E-01	1.15E-01	1.84E-01
575	3.65E-01	3.15E-01	4.17E-01	**575**	1.67E-01	1.31E-01	2.08E-01

APPENDIX K

DAYTIME HOURLY CONCENTRATION ESTIMATES AND MEASUREMENTS BY SEASON

This appendix relates to the evaluation described in section 3.2.5. This evaluation is intended to inform the extent to which occurrences of the relatively higher daytime concentration events at ambient air monitors are reflected in the distribution of daytime model predicted concentrations. The following steps were performed to prepare the ambient air monitor and modeled concentration data sets for this evaluation.

(1) *Selection of the datasets to best represent the distributions of ambient air monitor and model concentrations.* For the monitor data, we used the reported unadjusted (*as is*) values with no augmentation for time points where missing. The monitor datasets used met completeness criteria (section 3.5.1) for each year used (2011-2013), although some seasons may be relatively more (or less) complete than others, even within the 3-year pooled dataset. The AERMOD estimates provide a complete time-series of hourly ambient air concentrations in each year. Similar to the Appendix D evaluation, this evaluation uses the AERMOD estimates (*as is*) for each monitor site.

(2) *Stratification of the monitor and model distributions of hourly concentrations by time of day and season.* Time-of-day was split into two categories: daytime (hours most likely associated with population exposure) included the hours of 6AM to 8PM, and nighttime included all other hours. Seasons were stratified as winter (December, January, February), spring (March, April, May), summer (June, July, August), and fall (September, October, November). The result is 8 datasets of monitor concentrations and of model concentrations (winter daytime, winter nighttime, spring daytime, etc.) for each monitor location in each study area.

(3) *Addressing missing monitor concentrations for some time points (while having complete model concentrations).*[1] The model receptor concentration distribution was paired with the monitor concentration distribution based on percentile within each distribution.

The paired model and monitor concentration distributions were plotted for each of the two times-of-day and four seasons. Figure 3-1 presents the figures for the monitor location in each study area with the highest design value and for the daytime hours in the three warmer

[1] While the air quality model predicts values for every day and hour, an ambient air monitor typically does not measure for every hour in every year. Therefore, this distribution of values was calculated for each data set to have an equal reference point rather than compare only concentrations for reported measurements. To maximize the relative number of percentiles with respect to hours of data points in each season (which generally range from 2,730 to 3,864), the 0 to 100 percentiles were calcuated using every 0.04 percentile, thus 2,500 values were generated for every season and time of day pair.

months. The complete set of daytime graphs for all monitor locations and seasons are provided here.

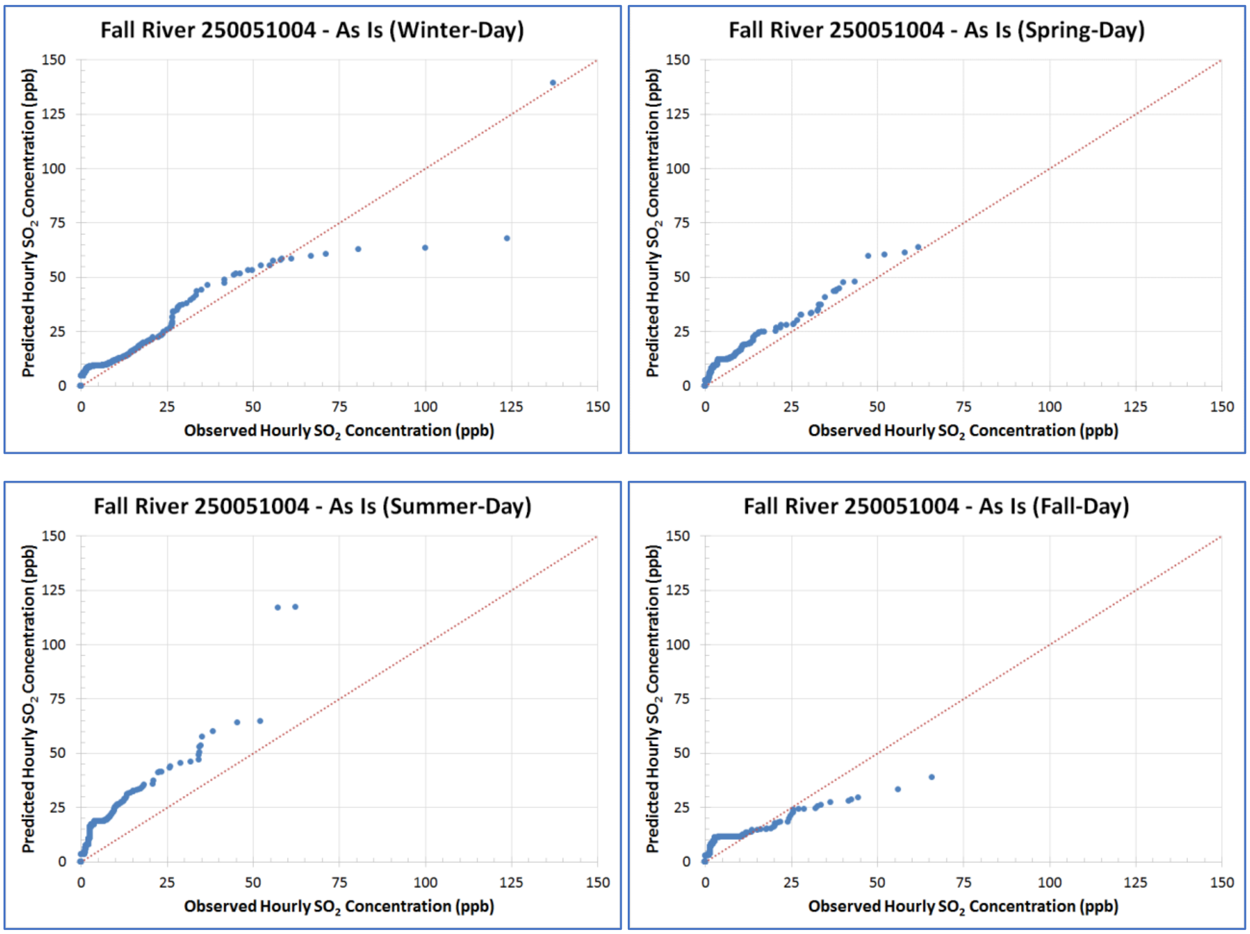


Figure K-1. Comparison of predicted and observed hourly SO_2 concentrations in ambient air (2011-2013): Fall River monitor 250051004 having the highest design value in study area.

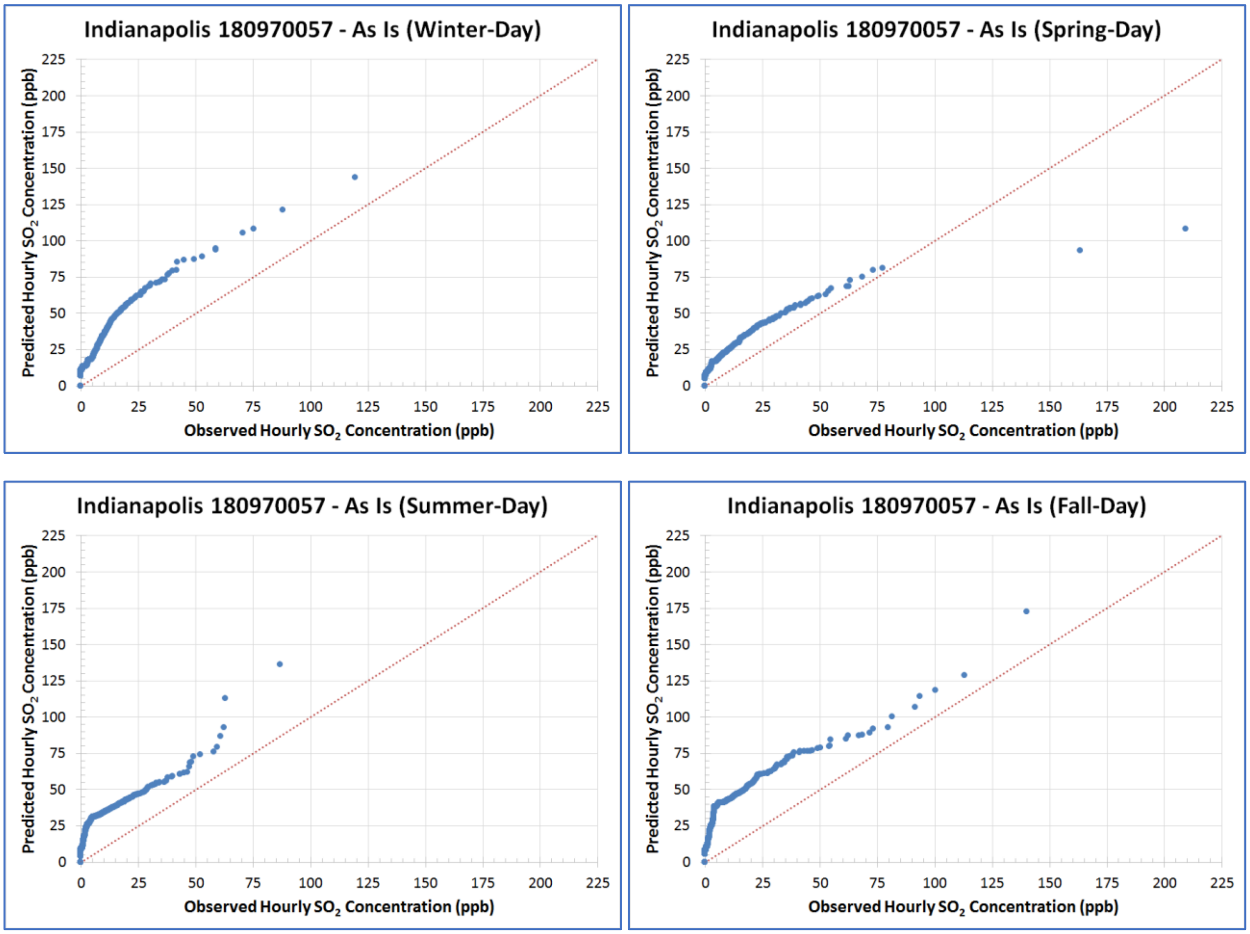


Figure K-2. Comparison of predicted and observed hourly SO_2 concentrations in ambient air (2011-2013): Indianapolis monitor 180970057 having the highest design value in study area.

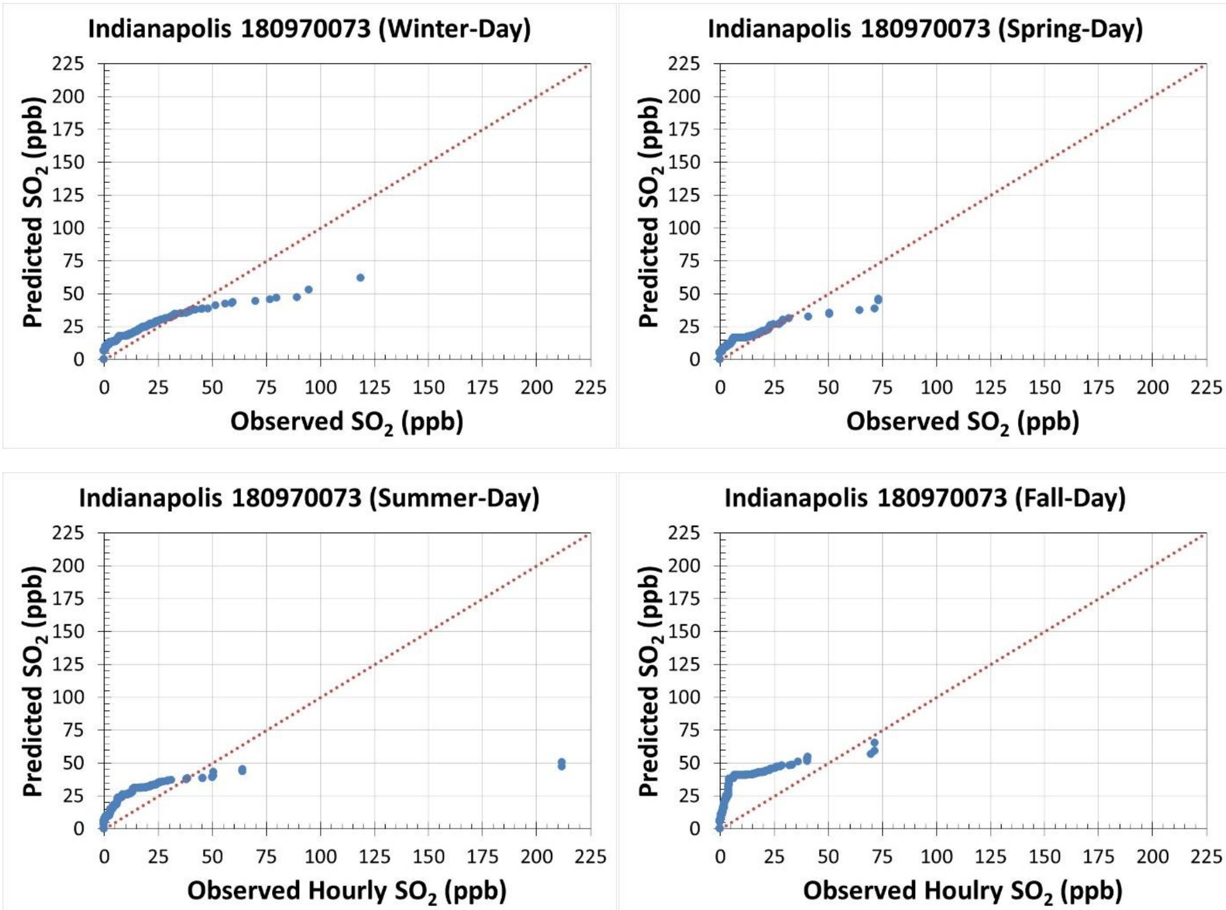


Figure K-3. Comparison of predicted and observed hourly SO_2 concentrations in ambient air (2011-2013): Indianapolis monitor 180970073.

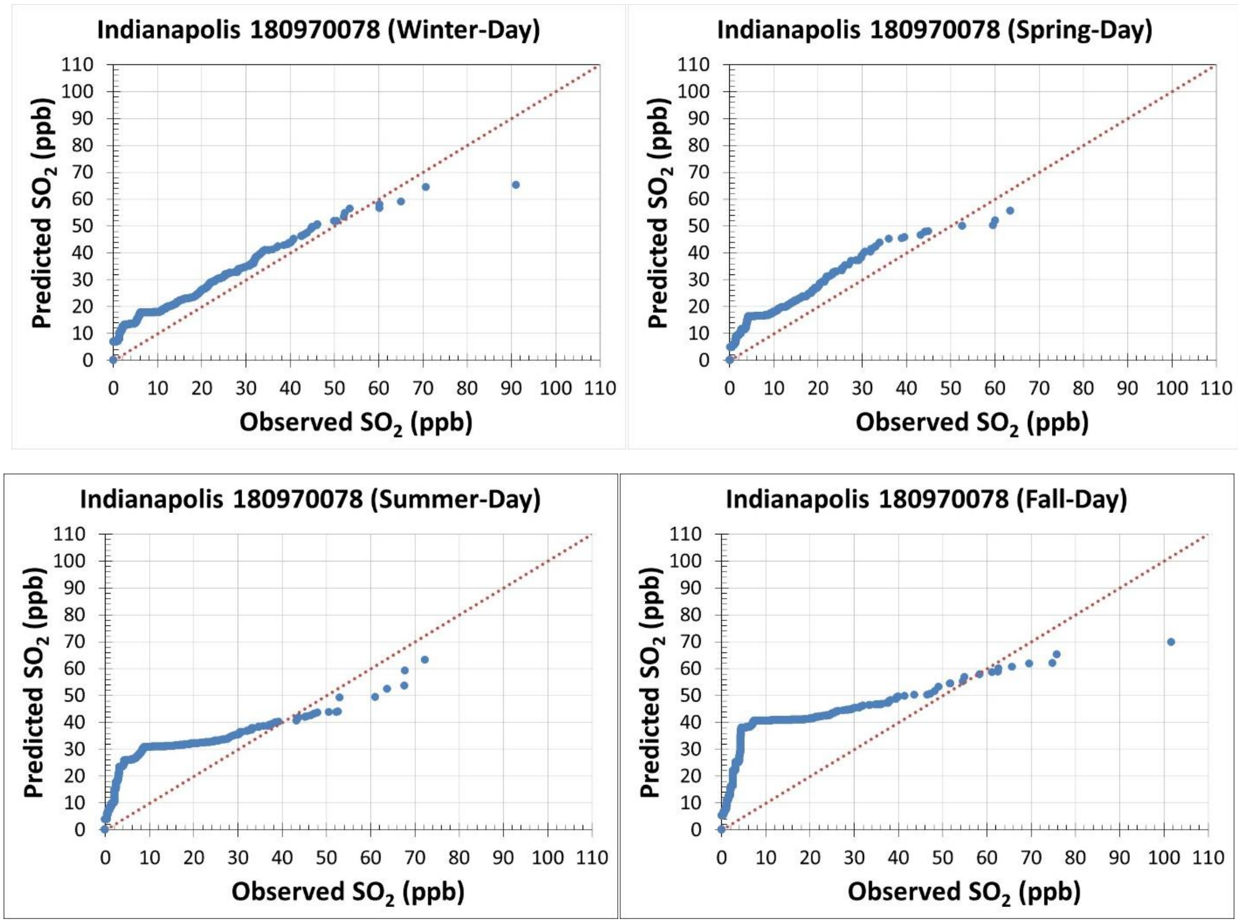


Figure K-4. Comparison of predicted and observed hourly SO_2 concentrations in ambient air (2011-2013): Indianapolis monitor 180970078.

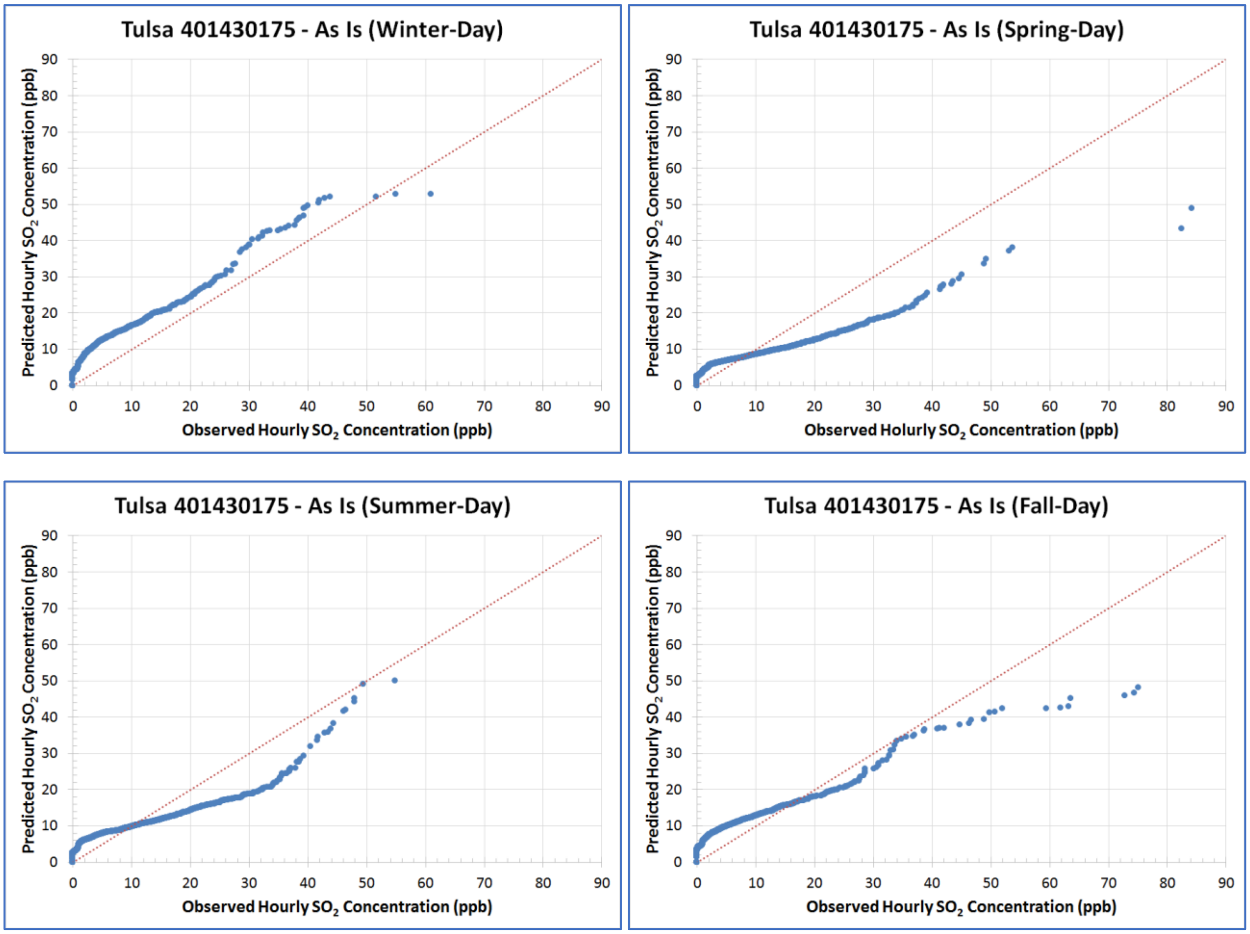


Figure K-5. Comparison of predicted and observed hourly SO_2 concentrations in ambient air (2011-2013): Tulsa monitor 401430175 having the highest design value in the study area.

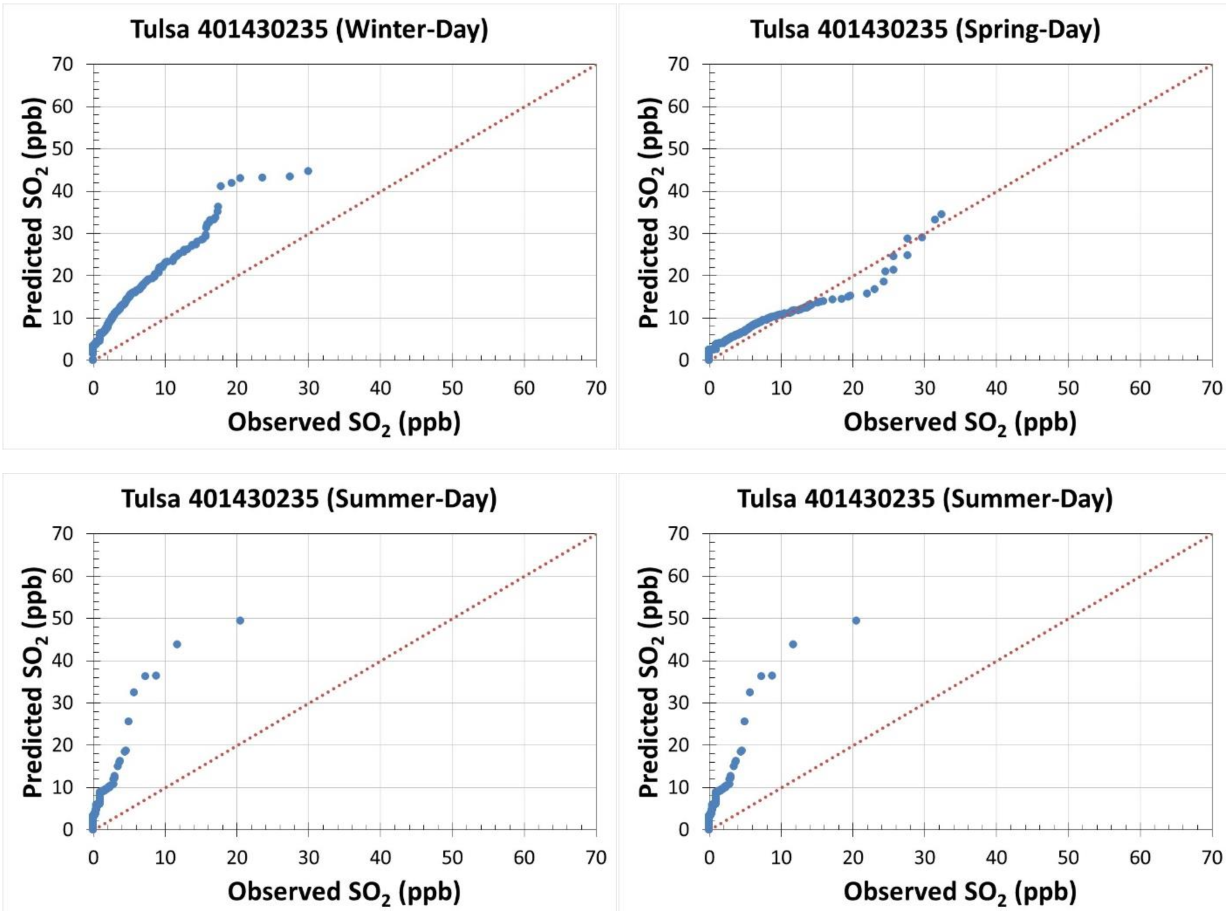


Figure K-6. Comparison of predicted and observed hourly SO_2 concentrations in ambient air (2011-2013): Tulsa monitor 401430235.

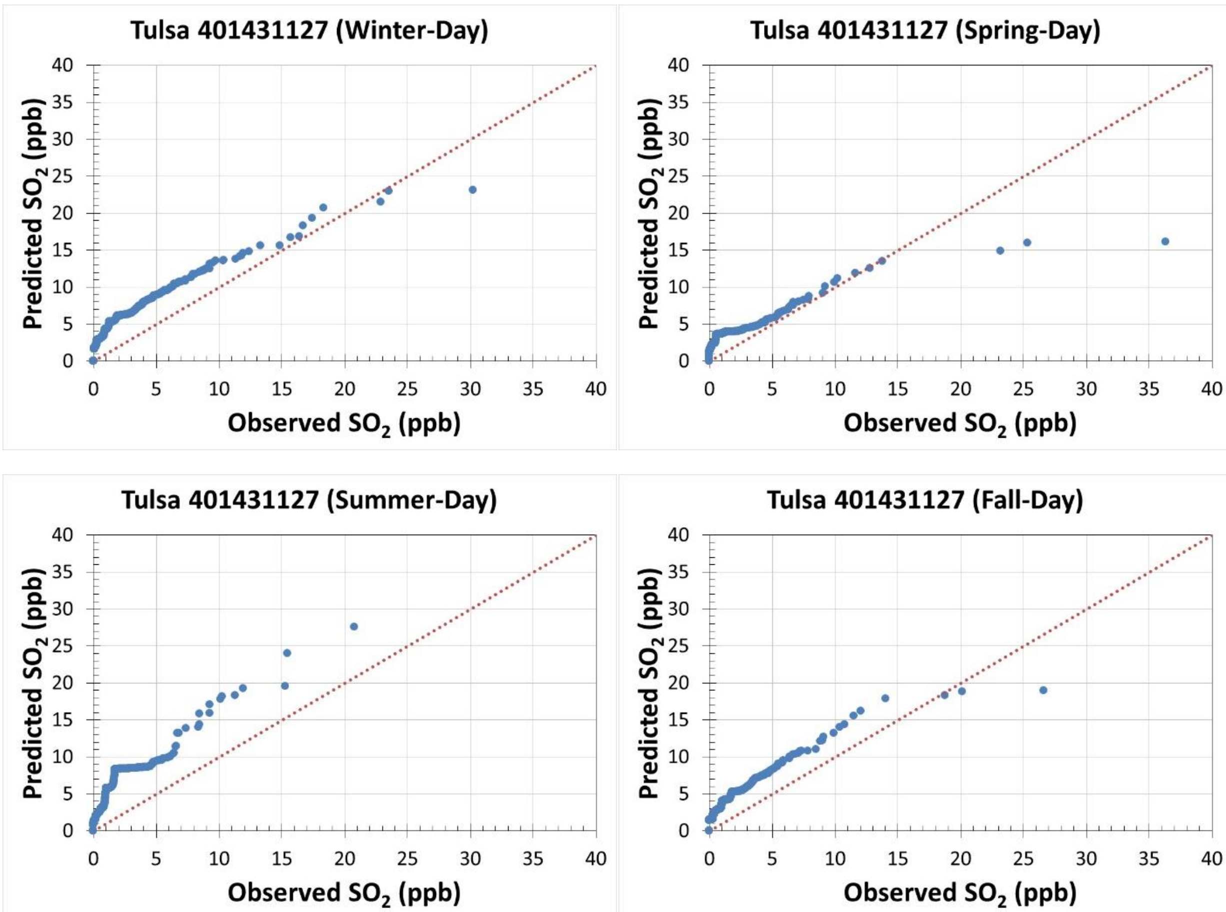


Figure K-7. Comparison of predicted and observed hourly SO_2 concentrations in ambient air (2011-2013): Tulsa monitor 401431127.

United States Environmental Protection Agency	Office of Air Quality Planning and Standards Health and Environmental Impacts Division Research Triangle Park, NC	Publication No. EPA-452/R-18-003 May 2018