Mortality Associated with Wildfire Smoke Exposure in Washington State, 2006-2017

Annie Doubleday

A thesis

submitted in partial fulfillment of the

requirements for the degree of

Master of Public Health

University of Washington

2019

Committee:

Tania Busch Isaksen

Lianne Sheppard

Ranil Dhammapala

Program Authorized to Offer Degree:

Environmental and Occupational Health Sciences

School of Public Health

University of Washington

Abstract

Mortality Associated with Wildfire Smoke Exposure in Washington State, 2006-2017

Annie Doubleday

Chair of the Supervisory Committee:

Tania Busch Isaksen

Department of Environmental and Occupational Health Sciences

Wildfire events are increasing in prevalence in the western United States. Research has found mixed results on the degree to which exposure to wildfire smoke increases the risk of mortality. A time-stratified case crossover study was employed to examine the odds of non-traumatic mortality on wildfire smoke days compared to non-wildfire smoke days in Washington State. Wildfire smoke exposure is characterized by daily average $PM_{2.5}$ concentrations on wildfire smoke days, from June 1 through September 30, for 2006-2017. Monitoring data was obtained from the Washington State Department of Ecology's monitoring network, and mortality data was obtained from the Washington State Department of Health. After adjusting for humidex, the odds of same-day non-traumatic mortality across the entire population was 1.0% (95% CI: 0.99 - 1.04) greater on wildfire smoke days compared to non-wildfire smoke days. Among adults ages 65-84, we estimated a 2.0% (95% CI: 0.98, 1.06) increase in the odds of same-day non-traumatic mortality, and an 8.0% (95% CI: 0.96, 1.21) increase in the odds of same-day respiratory mortality on wildfire smoke days compared to non-wildfire smoke days in this age group. Among all ages, we estimated a 1.3% (95% CI: 1.00, 1.02) increase in the odds of next-day non-traumatic mortality on wildfire smoke days compared to non-wildfire smoke days. This study is

the first to examine wildfire smoke and mortality in Washington State, and its findings are consistent with other research on wildfire smoke and mortality. This study lays the groundwork for estimating the risk associated with exposure to wildfire smoke in Washington, and will help inform local and state risk communication efforts and decision-making during future wildfire smoke events.

Table of Contents

List of Appendices

List of Tables

List of Figures

Abbreviations and Acronyms

95% CI: 95% confidence interval

AIRPACT: Air Indicator Report for Public Awareness and Community Tracking

AOD: Aerosol Optical Depth

CALPUFF: Modeling system simulating dispersion of atmospheric pollution

CO: Carbon monoxide

COPD: Chronic Obstructive Pulmonary Disease

CSN: Chemical Speciation Network

DOH: Washington State Department of Health

EC: Elemental Carbon

ECY: Washington State Department of Ecology

ED: Emergency Department

EPA: United States Environmental Protection Agency

FRP: Fire Radiative Power

HMS: NOAA Hazard Management System

ICD-10: 10th revision of the International Statistical Classification of Diseases and Related Health Problems

IMPROVE: Interagency Monitoring of Protected Visual Environments

IPCC: Intergovernmental Panel on Climate Change

MODIS: Moderate Resolution Imaging Spectroradiometer

NAAQS: National Ambient Air Quality Standards

NEHA: National Environmental Health Association

NO_2: Nitrogen dioxide

NOAA: National Oceanic and Atmospheric Administration

OC: Organic Carbon

OFM: Washington State Office of Financial Management

OR: Odds ratio

PM: Particulate matter

PM_{10}: Particulate matter less than 10 microns in diameter

$PM_{2.5}$: Particulate matter less than 2.5 microns in diameter

POM: Particulate organic matter

PSCAA: Puget Sound Clean Air Agency

RR: Relative risk

SD: Standard deviation

SMOKE: Sparse Matrix Operator Kernal (SMOKE) Modeling System

TEOM: Tapered element oscillating microbalance air quality monitor

WAQA: Washington Air Quality Advisory

WRF-Chem: Weather Research and Forecasting with Chemistry

WSEHA: Washington State Environmental Health Association

Acknowledgements

First and foremost, this project would not be what it is without the constant support of my advisor and committee chair, Tania Busch Isaksen. Tania spent countless hours guiding me at every step of the way, pushing me to always do better, and to seek out hard questions that are relevant to the practice community. This project would not have happened without her constant support, feedback and cheerleading, and I am extremely grateful for all the time, energy and effort she poured into this project. Lianne Sheppard, a committee member, provided technical support and guidance with methods development, and thoughtful review of my scientific writing. To her, I am grateful for pushing me to be a better scientist at every step of the way. And Ranil Dhammapala, my final committee member, provided his technical expertise, practice-based perspective, and insight that helped guide the project. All three committee members provided their insight and support on major aspects of this work, and spent countless hours in meetings and reviewing my materials.

My colleagues at the Washington State Department of Health and Department of Ecology provided additional technical support to ensure this project was a success. Jill Schulte (Ecology) provided hours of technical support and review and played a major role in the development of the exposure assessment work. Much of the credit for this work belongs to her. Matt Kadlec (Ecology) and Julie Fox (DOH) also supported the development of the project and reviewed several documents along the way, offering their insight and expertise.

Several staff in the Department of Environmental and Occupational Health Sciences supported efforts to help make this project a success. Brian High assisted with troubleshooting IT and data issues, Randy Katz assisted with scheduling and meeting logistics, and Trina Sterry and Christine Tran provided administrative support.

Eddie Kasner, a post-doc in DEOHS, and Miriam Calkins, a recent doctoral student in DEOHS, provided technical expertise regarding study design choices, proving to be crucial to the development of my methods and overall project. Additionally, students and faculty from the UW Biostatistics Consulting group assisted with model and design choices.

My DEOHS cohort, family and friends supported me throughout the entire process, and believed in me when I did not. Thank you to Megumi Matsushita, Claire Pendergrast, the Borg-Doubleday family and so many others for your constant support.

Literature Review

The western United States is seeing an increase in wildfires during the summer and fall months, elevating the importance of understanding the health impacts associated with exposure to wildfire smoke (Dennison, Brewer, Arnold, & Moritz, 2014; Westerling, 2016). It is estimated that the total forest fire area burned from 1984 to 2015 was nearly twice as much as expected compared to a scenario without climate change (Abatzoglou & Williams, 2016). This trend is expected to worsen, with climate projections indicating wildfires in the western U.S. will increase in frequency and intensity (Joyce et al., 2014; Littell, McKenzie, Peterson, & Westerling, 2009). The Intergovernmental Panel on Climate Change (IPCC) estimates that climate change will increase the wildfire season in North America by 10-30% (Parry, 2007). This is expected to result in worsening air quality due during wildfire season in the coming decades (Interagency Working Group on Climate Change and Health, 2010).

Wildfire smoke contains a wide range of compounds known to be harmful to human health, including fine particulate matter ($PM_{2.5}$), acrolein, benzene, carbon monoxide, and polycyclic aromatic hydrocarbons (Naeher et al., 2010; Urbanski, Hao, & Baker, 2008). Exposure to these toxic compounds is of concern near the source, as well as several hundred miles away (Moeltner, Kim, Zhu, & Yang, 2013; Munoz-Alpizar et al., 2016). While it has been shown that the toxic compounds from wildfire smoke travel long distances from the source, potentially exposing thousands of individuals, the health effects associated with exposure are just beginning to be understood (Black, Tesfaigzi, Bassein, & Miller, 2017).

Evidence for an association between exposure to wildfire smoke and all-age non-traumatic mortality is mixed. A handful of studies provide evidence for an association between short-term exposure to wildfire smoke and increased risk of mortality for all ages (Haikerwal et al., 2015; Hänninen et al., 2009; F. Johnston, Hanigan, Henderson, Morgan, & Bowman, 2011; Linares, Carmona, Tobías, Mirón, & Díaz, 2015; Vila, Nunes, Ignotti, & Hacon, 2013), while others do not (Faustini et al., 2015; Hong, King, Saraswat, & Henderson, 2017; F. Johnston et al., 2011; Kollanus et al., 2017). One study in Finland reported a 1.0% (95% CI: 0.8 – 2.1%) increase in all-cause mortality per 10 $\mu g/m^3$ increase in same-day $PM_{2.5}$ concentration (Hänninen et al., 2009), while another study in Australia estimated that wildfire events were associated with a 2.0% (95% CI: -2.0 – 5.0%) increase in the odds of same-day non-accidental mortality (F. Johnston et al., 2011). A third study, pooling effect estimates from 10 European cities reported a 1.78% (95% CI: -0.91 – 4.53%) increase in combined same-day and previous day natural mortality on smoke-affected days (Faustini et al., 2015). A final study in Finland estimate a 0.2% (95% CI: -7.6 – 8.6%) increase in non-accidental mortality for all ages for a 10 $\mu g/m^3$ increase in same-day $PM_{2.5}$ concentration on smoke-affected days (Kollanus, Tiittanen, Niemi, & Lanki, 2016).

Several studies also report estimates for respiratory and cardiovascular mortality, reporting some or no evidence of association with respiratory mortality (Faustini et al., 2015; F. Johnston et al., 2011; Morgan et al., 2010) and small increases in the odds of cardiovascular mortality (Faustini et al., 2015; F. Johnston et al., 2011). Faustini et al. observe a 0.65% (95% CI: 0.10 – 1.19%: lag 0-5) increase in cardiovascular mortality per 10$\mu g/m^3$ increase in PM_{10} concentration for the five days prior to death, and a 2.13% (95% CI: 0.85 – 3.42%) increase in respiratory mortality per 10$\mu g/m^3$ increase in PM_{10} concentration for the five days prior to death (Faustini et al., 2015). Further, Johnston et al. estimate a 6.0% (95% CI: 0.97 – 1.17) increase in the odds of same-day cardiovascular mortality and a 0% (95% CI: 0.79 – 1.25) increase in the odds of same-day respiratory mortality, on wildfire smoke days compared to non-wildfire smoke days (F. Johnston et al., 2011). Finally, Morgan et al. report a 0.76% (95% CI: -0.76 – 2.30%) increase in same-day all age cardiovascular mortality, and a -0.32% (95% CI: -3.70 – 3.18%) change in same-day all age respiratory mortality per 10μg/m^3 increase in bushfire PM_{10} concentration (Morgan et al., 2010).

A subset of these studies has reported significant associations between wildfire smoke exposure and non-traumatic mortality in groups 65 years and older, but less of an effect in other age groups (Analitis, Georgiadis, & Katsouyanni, 2012; Haikerwal et al., 2015; Shaposhnikov et al., 2014; Vila et al., 2013). Analitis et al. (2011) reports that the effect of respiratory mortality in Greece was greater in adults 75 and over for large fires (Analitis et al., 2012). Further, Haikerwal et al. (2015) observed an increase in risk of cardiac arrests, especially in adults 65 and over in Australia, and Vila et al. (2013) reports that adults 65 and over in their study in Brazil have the strongest association between exposure to biomass burning and circulatory disease mortality (Haikerwal et al., 2015; Vila et al., 2013). There are few U.S.-based mortality studies, providing little evidence for U.S.-specific mortality associated with wildfire smoke exposure. Further, there are no mortality studies in the Pacific Northwest, necessitating further research on the association between wildfire smoke exposure and non-traumatic mortality in Washington State.

Decades of research has led to a weight of evidence inference that there is likely a causal relationship between exposure to anthropogenic $PM_{2.5}$ and increased morbidity, including hospitalizations and ED visits for asthma and COPD, among other illnesses (Li et al., 2016; Orellano, Quaranta, Reynoso, Balbi, & Vasquez, 2017; US EPA, 2009). Only recently has research on wildfire smoke begun to examine similar endpoints. Studies in the western U.S., Australia, and Europe have repeatedly pointed towards an association between exposure to wildfire smoke and various respiratory outcomes, including asthma and COPD, typically measured by hospitalizations or ED visits (Cascio, 2018; C E Reid et al., 2016). This association is seen even more commonly among older adults, typically defined as adults 65 and over (Delfino et al., 2009; Morgan et al., 2010; Mott et al., 2005; Wettstein et al., 2018). A handful of studies have reported that adults 65 years and over have the greatest increase in all respiratory hospitalizations and readmissions (Morgan et al., 2010; Mott et al., 2005; Wettstein et al., 2018). Some have reported increases in specific respiratory outcomes among older adults, including increased risk of COPD or asthma compared to other age groups (Delfino et al., 2009; Morgan et al., 2010; Mott et al., 2005; Wettstein et al., 2018). Further, although many studies have looked at the potential association between wildfire smoke exposure and various cardiovascular outcomes, results are mixed, indicating no clear association, and very little evidence for an association among older adults (C E Reid et al., 2016). One study in particular found an increased risk of all-cause cardiovascular ED visits, especially among those 65 years and older (Wettstein et al., 2018), while another study found an increase in all-cause cardiac ED visits for adults with a 10 $\mu g/m^3$ increase in 24-hour $PM_{2.5}$ (Tinling, West, Cascio, Kilaru, & Rappold, 2016). However, the majority of studies examining the association between exposure to wildfire smoke and cardiovascular morbidity find little evidence for a positive association (Alman et al., 2016; Delfino et al., 2009; C E Reid et al., 2016).

An important challenge in wildfire smoke epidemiology is the lack of a gold standard method for defining what constitutes a wildfire smoke affected day. Common methods utilize air quality area monitoring particulate matter (PM) measurements (Alman et al., 2016; Dennekamp et al., 2015; Haikerwal et al., 2015; Hänninen et al., 2009; Hong et al., 2017; F. Johnston et al., 2011; Linares, Carmona, Salvador, & Díaz, 2018; Linares et al., 2015; Sastry, 2002; Shaposhnikov et al., 2014; Zu, Tao, Long, Goodman, & Valberg, 2016), satellite data (Delfino et al., 2009; Elliott, Henderson, & Wan, 2013; Faustini et al., 2015; Wettstein et al., 2018; Yao, Eyamie, & Henderson, 2016), air pollution chemistry modeling (Gan et al., 2017; Grell et al., 2005; Jia Coco Liu et al., 2017), or a combination of methods in an attempt to create a more robust definition (Dell, Ford, Fischer, & Pierce, 2019; Gan et al., 2017; Henderson et al., 2011; Kollanus et al., 2016; Martin, Hanigan, Morgan, Henderson, & Johnston, 2013; Yao & Henderson, 2014), described in detail below.

Exposure to wildfire smoke is most commonly assessed through PM monitoring data, typically PM less than 2.5 microns in diameter ($PM_{2.5}$) or less than 10 microns (PM_{10}) (Alman et al., 2016; Dennekamp et al., 2015; Haikerwal et al., 2015; Hänninen et al., 2009; Hong et al., 2017; F. Johnston et al., 2011; Linares et al., 2018, 2015; Sastry, 2002; Shaposhnikov et al., 2014; Zu et al., 2016). Within PM monitoring data, some studies examine the health impact for each $10\mu g/m^3$ increase in $PM_{2.5}$ (Alman et al., 2016; Hänninen et al., 2009), while other studies examine the impact above a specific threshold, such as the 99th percentile of $PM_{2.5}$ for the time series (F. Johnston et al., 2011; Martin et al., 2013). Use of PM monitoring data, while relatively straightforward to implement due to the ubiquity and availability of $PM_{2.5}$ monitoring data, may incorrectly attribute days with high anthropogenic particulate matter as wildfire smoke days. Use of PM monitoring data is also subject to exposure misclassification due to inadequate spatial coverage across the area of study, and relies on area monitors to serve as a proxy for individual exposure.

Another method of determining wildfire smoke exposure is to use Moderate Resolution Imaging Spectroradiometer (MODIS) satellite imagery data, such as Aerosol Optical Depth (AOD) or Fire Radiative Power (FRP) to serve as a proxy for PM or to determine smoke days based on the distribution of AOD or FRP over the time series (Delfino et al., 2009; Elliott et al., 2013; Faustini et al., 2015; Yao et al., 2016). Additionally, MODIS imagery is used to generate the NOAA Hazard Management System (HMS) fire and smoke product, to serve as a binary indicator of smoke presence on a daily basis (National Oceanic and Atmospheric Administration, 2018; Wettstein et al., 2018; Yao & Henderson, 2014). These methods are often used when reliable and adequate air monitoring measurements are unobtainable. However, as with area monitoring-based methods, satellite-based methods are prone to misclassification. With satellite data, the height of the smoke plume is typically unknown, and therefore, exposure on the ground may be very different from what is observed based on the AOD or FRP measurements. This technology, however, is rapidly improving, and may serve as a promising method for future research.

A third common method is to simulate daily $PM_{2.5}$ concentrations using the Weather Research and Forecasting with Chemistry (WRF-Chem) model (Grell et al., 2005). The WRF-Chem model combines meteorological conditions with atmospheric chemical processes to simulate air quality, and daily $PM_{2.5}$ concentrations. Studies have shown that this model simulates estimates that are, on average, biased higher compared to observations, and yield differing estimates of smoke-attributed days compared to other methods (Gan et al., 2017; Grell et al., 2005; Jia Coco Liu et al., 2017).

Several studies combine the above methods to increase the accuracy of their wildfire smoke day definition, or to validate their primary method of assessment (Dell et al., 2019; Gan et al., 2017; Henderson et al., 2011; Kollanus et al., 2016; Martin et al., 2013; Yao & Henderson, 2014). However, because no gold standard method exists, it is not clear that combining methods necessarily yields more accurate estimation of wildfire smoke attributable days. Further, many of the differences in effect estimates across studies may be due to the utilization of various exposure assessment methodologies, which cannot be directly compared (Gan et al., 2017). More research in determining wildfire smoke affected days needs to be conducted to validate methods currently in use, and for more accurate comparison of effect estimates across studies.

Only two efforts have examined health effects of wildfire smoke in Washington State, both focused on the extreme fire season of 2012 and with limited geographical coverage to Central Washington (Gan et al., 2017; Washington State Department of Health, 2012). A group of local health departments in Central Washington in conjunction with the state Department of Health examined cardiopulmonary emergency department (ED) and outpatient visits in the affected areas before and after a major wildfire smoke event, and found a substantial increase after the wildfire smoke event compared to before. They found a 28% (95% CI: 18 – 39%) increase for cardiovascular disease and respiratory outpatient visits in the two weeks during the wildfires compared to the two weeks before, and an 18% (95% CI: 9 – 28%) increase in the two weeks after the wildfires compared to the two weeks before. While providing evidence for health effects from wildfire smoke exposure in Washington, this study was limited by its small sample size, and did not examine the possible effect on mortality (Washington State Department of Health, 2012). Gan et al. conducted a time-stratified case-crossover study of the same event, and found a 7.6% (95% CI: 1.9 – 13.6%) increase in the risk of cardiopulmonary hospitalizations for each 10 $\mu g/m^3$ increase in wildfire smoke $PM_{2.5}$ using exposure based on geographically weighted ridge regression that includes input from the WRF-Chem model, satellite observed aerosol optical depth and kriged $PM_{2.5}$ (Gan et al., 2017). While both of these studies offer insight into the health effects of wildfire smoke exposure in Washington, they are limited in duration and sample size, and only examine hospitalizations. No studies have been conducted on the potential risk of mortality associated with wildfire smoke exposure in Washington, with the closest mortality study conducted in British Columbia in 2017 (Hong et al., 2017). Our study examines the association between wildfire smoke exposure and the odds of non-traumatic mortality in Washington State over 12 years using an exposure assessment method that reduces misclassification. We hypothesized *a priori* that evidence of an elevated risk would be limited to individuals 65 and over, to an effect at a lag of one or more days, and within respiratory and cardiovascular causes, consistent with many of the effects seen in the literature discussed above.

Study Aims

Wildfire events are increasing in prevalence in the western United States (Dennison et al., 2014; Westerling, 2016). While there is no consistent evidence for an association between exposure to wildfire smoke and non-traumatic mortality, there is some evidence for an association among those 65 years and older (Cascio, 2018; Haikerwal et al., 2015; Vila et al., 2013; Youssouf et al., 2014). Further research is needed to support the potential association between wildfire smoke exposure and mortality among older adults, specifically in Washington State, where little research has been done on the health effects of wildfire smoke exposure (Jia C Liu, Pereira, Uhl, Bravo, & Bell, 2015; C E Reid et al., 2016).

The overall objective is to examine the potential association between daily average $PM_{2.5}$ exposure on wildfire smoke days and non-traumatic mortality in Washington State. Our central hypothesis is that we will observe no change in the odds of all same-day non-traumatic mortality, and a small increase in the odds of same-day cardiovascular and respiratory mortality on wildfire smoke days compared to non-wildfire smoke days. Further, we hypothesize that we will observe an increase in the odds of all same-day non-traumatic mortality for adults 65 and over, and an increase in the odds of all next day non-traumatic mortality, given wildfire smoke exposure on the prior day compared to no wildfire smoke exposure. The overall rationale for this work is to help focus further research regarding the health effects of wildfire smoke exposure in Washington, and to better target intervention and communication strategies in the state. This will be achieved through a time-stratified case-crossover study examining non-traumatic deaths in Washington and wildfire smoke exposure.

Aim 1: Determination of a wildfire smoke day definition and misclassification of other definitions

1.1 Determine wildfire smoke day definitions for Washington State.

1.2 Compare wildfire smoke day definitions using organic carbon chemical speciation data.

1.3 Quantitatively compare wildfire smoke affected day definitions to definitions from a study conducted in Washington State by Gan et al.

Aim 2: Evaluate the association between wildfire smoke exposure and non-traumatic mortality through a case-crossover study design.

2.1. Evaluate the association between wildfire smoke exposure and all same-day non-traumatic mortality, by cause of death, by age group, and due to previous day exposure

2.2. Stratify results by race, age group, sex, socioeconomic status, cause of death and location of death

Aim 3: Communicate findings to partner agencies, conferences, and to the public.

3.1. Draft research findings for submission to scientific journals

3.2. Present findings to professional environmental health associations (WSEHA, NEHA)

3.3. Present findings to state and local health and environmental agencies (DOH, ECY, PSCAA, EPA)

These aims will positively impact the field by adding evidence to the potential association between wildfire smoke exposure and non-traumatic mortality, and will help fill the gap in the literature around the health impacts of wildfire smoke exposure in Washington state (Gan et al., 2017; Jia C Liu et al., 2015; C E Reid et al., 2016).

Aim 1: Classifying Wildfire Smoke Definitions

Classification of Wildfire Smoke Definitions using Organic Carbon Speciation data

Annie Doubleday[1], Jill Schulte[2], Ranil Dhammapala[2], Lianne Sheppard[1,3], Tania Busch Isaksen[1]

[1]Department of Environmental and Occupational Health Sciences, University of Washington
[2]Air Quality Program, Washington State Department of Ecology
[3]Department of Biostatistics, University of Washington

Abstract

Many epidemiological studies of the health effects of wildfire smoke exposure depend on accurate classification of days as either smoke affected or not. No current studies quantify classification agreement of wildfire smoke exposure assessment approaches. Four definitions for classifying wildfire smoke-affected days developed using Washington State exposure data were compared. The classification agreement of each definition with organic carbon concentrations from Washington Chemical Speciation and IMPROVE Network data was calculated. The four classification definitions were then compared to three approaches from the Gan et al. study based in Washington State. The wildfire smoke classification definition that uses a set of criteria to categorize days performs the best across nearly all metrics. Further research is needed to develop a gold standard for assessing wildfire smoke exposure to improve risk estimates of associated health outcomes, and for more accurate comparison of effect estimates across studies.

Background

Wildfires are increasing in frequency and severity during the summer and fall months in the western United States (Westerling, 2016). Efforts to quantify the risk of several health endpoints from exposure to wildfire smoke have increased significantly over the last decade, including mortality, hospitalizations and ED visits (Black et al., 2017; Cascio, 2018; Youssouf et al., 2014). However, there is no common approach for classifying wildfire smoke-affected days or time units, and to our knowledge, there are only two studies that attempt to compare approaches (Gan et al., 2017; Henderson et al., 2011). These two studies, however, do not quantify the classification agreement of days (Gan et al., 2017; Henderson et al., 2011). The field would benefit from a comparison of classification definitions, as well as development of a gold standard to support accurate comparisons of risk across studies. In this paper, we consider several different approaches for classifying wildfire smoke days using exposure data from Washington State from 2006-2017, and quantify their classification agreement using: 1) organic carbon data from the Chemical Speciation and Interagency Monitoring of Protected Visual Environments (IMPROVE) Networks, and 2) wildfire smoke $PM_{2.5}$ definitions from Gan et al., the only other wildfire smoke and health effects study conducted in Washington State (Gan et al., 2017).

Assessing exposure from wildfire smoke is challenging, because no gold standard approach for defining a 'wildfire smoke affected day' exists in the exposure assessment literature. Common data sources include utilization of area air quality monitoring particulate matter (PM) measurements, satellite data, air pollution chemistry modeling, or a combination of data sources to create a more robust definition.

The first, and most commonly used, data source to assess exposure to wildfire smoke is through PM monitoring data, typically PM less than 2.5 microns in diameter ($PM_{2.5}$) or less than

10 microns (PM_{10}) (Alman et al., 2016; Dennekamp et al., 2015; Hänninen et al., 2009; Hong et al., 2017; F. Johnston et al., 2011; Martin et al., 2013; Sastry, 2002; Shaposhnikov et al., 2014). Within this data source, some studies do not distinguish wildfire smoke $PM_{2.5}$ from other sources, and model smoke continuously, examining the health impact for each 10μg/m^3 increase in $PM_{2.5}$ (Alman et al., 2016; Hänninen et al., 2009). Others examine the impact above a specific percentile of the $PM_{2.5}$ distribution, for example the 99th percentile of $PM_{2.5}$ for the time series (F. Johnston et al., 2011; Martin et al., 2013), or the impact above a specific $PM_{2.5}$ concentration identified *a priori* (Dennekamp et al., 2015; Sastry, 2002). Use of $PM_{2.5}$ monitoring data, while relatively straightforward to implement due to the ubiquity and availability of these data, may incorrectly attribute days with high anthropogenic particulate matter as wildfire smoke days, and thus may attribute some of the associated health effect estimates to elevated anthropogenic $PM_{2.5}$. In some instances, this may be suitable, depending on the scientific question of interest. However, for studies interested in separating wildfire smoke $PM_{2.5}$ and background anthropogenic $PM_{2.5}$, this produces exposure misclassification. Further, use of $PM_{2.5}$ monitoring data is subject to additional exposure misclassification, as it relies on area monitors to serve as a proxy for individual exposure, and often does not provide adequate spatial coverage for the full population at risk.

The second data source used for determining wildfire smoke exposure is to use metrics derived from Moderate Resolution Imaging Spectroradiometer (MODIS) satellite imagery data, such as Aerosol Optical Depth (AOD) or Fire Radiative Power (FRP) to serve as a proxy for PM or to determine wildfire smoke days based on the distribution of AOD or FRP over the time series (Delfino et al., 2009; Elliott et al., 2013; Faustini et al., 2015; Yao et al., 2016). Additionally, MODIS imagery is used to generate the NOAA Hazard Management System

(HMS) fire and smoke product, to serve as a binary indicator of smoke presence on a daily basis (National Oceanic and Atmospheric Administration, 2018; Wettstein et al., 2018; Yao & Henderson, 2014). Satellite data are often used when reliable and adequate air monitoring measurements are unobtainable. However, as with area monitoring, satellite measurements are prone to misclassification. With satellite data, the height of the smoke plume is typically unknown, and therefore, exposure on the ground may be very different from what is observed based on the AOD or FRP measurements. This technology, however, is rapidly improving, and may serve as a promising method for future research.

The final commonly used source of data is to simulate daily $PM_{2.5}$ concentrations using models such as the Weather Research and Forecasting with Chemistry (WRF-Chem) (Grell et al., 2005). The WRF-Chem model combines meteorological conditions with atmospheric chemical processes to simulate air quality, and daily $PM_{2.5}$ concentrations. Studies have shown that this model simulates estimates that are, on average, biased higher compared to observations, and yield differing estimates of wildfire smoke-attributed days compared to other methods (Gan et al., 2017; Grell et al., 2005; Jia Coco Liu et al., 2017).

Several studies combine the above data sources to increase the accuracy of their wildfire smoke day definition, or to validate their primary approach (Dell et al., 2019; Henderson et al., 2011; Kollanus et al., 2016; Martin et al., 2013; Yao & Henderson, 2014). However, because no gold standard data source and classification exist, it is not clear that combining data sources necessarily yields more accurate estimation of wildfire smoke attributable days.

An additional approach to measuring wildfire smoke intensity has been posited through use of chemical speciation data. Several studies have explored the relationship between chemical speciation data and wildfire smoke, and have found organic carbon, and the elemental carbon to organic carbon ratio, to be good indicators of wildfire smoke (Ames, Fox, Malm, Schichtel, & Service, n.d.; Jaffe, Hafner, Chand, Westerling, & Spracklen, 2008; Kang, Gold, & Koutrakis, 2014; Jia C Liu & Peng, 2018; Mallia, Lin, Urbanksi, Ehleringer, & Nehrkorn, 2014; Spracklen et al., 2007; Vicente et al., 2013). Organic carbon is a major component of $PM_{2.5}$ from wildfire smoke (Jaffe et al., 2008), and one study finds that summertime organic carbon (OC) in the western U.S. is primarily due to smoke from wildfires (Spracklen et al., 2007). Three studies report organic carbon as a percent of $PM_{2.5}$ during wildfire smoke events. In one study in Portugal, Alves et al. found organic carbon makes up 31 to 55% of $PM_{2.5}$ by mass (Alves et al., 2011) and in another study in Portugal, Vicente et al. found organic carbon makes up 51 ± 12% of $PM_{2.5}$ by mass (Vicente et al., 2013). A third study in the western U.S. reported 51% as the median percentage of organic carbon in total $PM_{2.5}$ on wildfire smoke days (Jia C Liu & Peng, 2018). Further, Hand et al. examined IMPROVE and CSN data to approximate seasonal background particulate matter concentrations for U.S. regions. Regional background averages were calculated based on interpolated particulate matter concentration from a kriging model. The background concentration for particulate organic matter (POM) for the western US was 2-3 $\mu g/m^3$ (Hand et al., 2013), calculated as 1.8*OC, based on published measurements of organic carbon as a proportion of particulate organic matter (Malm, Schichtel, & Pitchford, 2011; Turpin, Lim, Turpin, & Lim, 2010). Other studies have proposed using the elemental (EC) to organic carbon ratio as a proxy for wildfire smoke intensity (Hand et al., 2013; Jaffe et al., 2008; Spracklen et al., 2007), as it provides information about the source contributions. While wildfire smoke contributes to both EC and OC, wildfire smoke makes up a larger proportion of total OC (Spracklen et al., 2007). Spracklen et al. report greater interannual variability of OC measurements compared to EC, pointing to wildfire smoke as main contributor to variability in seasonal OC (Spracklen et al., 2007).

To the author's knowledge, only two efforts have attempted to compare exposure assessment methods for classifying wildfire smoke affected days (Gan et al., 2017; Henderson et al., 2011). The first study, based in Washington, employed three different approaches for determining smoke affected $PM_{2.5}$. In the first approach, measured $PM_{2.5}$ from air quality monitors was interpolated across the state to a 15x15 km grid using ordinary kriging. In the second approach, daily $PM_{2.5}$ concentrations were simulated to a 15x15 km grid using the Weather Research and Forecasting with Chemistry (WRF-Chem) model. The third approach used geographically-weighted ridge regression to estimate $PM_{2.5}$ concentrations at each grid by combining aerosol optical depth (AOD) data, kriged concentrations, and WRF-Chem estimates (Gan et al., 2017; Lassman, Ford, Gan, Fischer, & Pierce, 2017). For each approach, they categorized days greater than 10 $\mu g/m^3$ as wildfire smoke affected (Gan et al., 2017). The second study, led by Henderson et al., also compared three approaches, using PM_{10} measurements as the basis for their exposure assessment. The first approach used PM_{10} measurements from six air quality monitors in British Columbia, the second approach estimated smoke PM_{10} from a CALPUFF dispersion model, originally developed by the California Air Resources Board, and the third approach used a SMOKE model developed by NOAA to identify smoke plumes based on satellite imagery (Henderson et al., 2011). Neither study compares relative classification across data sources and approaches. This study compares the relative classification across four different approaches for classifying wildfire smoke affected days in Washington State using organic carbon data from the Chemical Speciation and IMPROVE Networks and wildfire smoke methods from Gan et al. (Gan et al., 2017).

Methods

PubMed and Web of Science were searched for epidemiological literature on the health impacts of wildfire smoke conducted over the last 15 years and the methods for defining wildfire smoke affected days were documented for 30 studies. Key terms included "wildfire smoke", "wildland fire", "mortality", "hospitalizations", and "health effects". The majority of studies used a $PM_{2.5}$ threshold, above which was considered wildfire smoke affected (Alman et al., 2016; Dennekamp et al., 2015; Haikerwal et al., 2015; Hänninen et al., 2009; Hong et al., 2017; F. H. Johnston et al., 2014; Linares et al., 2018; Shaposhnikov et al., 2014), while others used data from satellite imagery (Delfino et al., 2009; Elliott et al., 2013; Faustini et al., 2015; Yao et al., 2016), the Weather Research and Forecasting with Chemistry (WRF-Chem) model or related models (Grell et al., 2005), or a combination of data sources (Dell et al., 2019; Henderson et al., 2011; Yao & Henderson, 2014). Table 1 displays a summary of the common wildfire smoke definitions in the literature. In our development of wildfire smoke definitions, we focused on approaches utilizing $PM_{2.5}$ from area monitors to define wildfire smoke affected days, given the ubiquity and ease of access of daily $PM_{2.5}$ data, and myriad examples in the literature.

Table 1. Wildfire Smoke Definitions in the literature

Data type	Inputs	Description	Associated studies
$PM_{2.5}$ monitoring data			
Percentile-based threshold	24-hr average $PM_{2.5}$ concentration	Smoke days categorized as those above a specific percentile of the distribution in the time series	Johnston et al. 2011 Shaposhnikov et al. 2014 Martin et al. 2013
Concentration-based threshold	24-hr average $PM_{2.5}$ concentration	Smoke days categorized above a specific concentration	Dennekamp et al. 2015 (plus additional criteria) Sastry 2002
Unit increase in $PM_{2.5}$ or PM_{10}	24-hr average $PM_{2.5}$ concentration	Impact of smoke measured for a 10	Alman et al. 2016 Hänninen et al. 2009

		unit increase in $PM_{2.5}$ or PM_{10}	Hong et al. 2017
$PM_{2.5}$ used as basis of prediction model	24-hr average $PM_{2.5}$ concentration	Measured $PM_{2.5}$ used as the basis for a prediction model (based on kriging or inverse distance weighting)	Gan et al. 2017 Delfino et al. 2009
Satellite-based data			
MODIS imaging	Includes AOD; FRP; and NOAA-based smoke plumes	MODIS – based data supplemented with interpolated $PM_{2.5}$	Delfino et al. 2009 Elliott et al. 2013 Faustini et al. 2015 Yao et al. 2016 Gan et al. 2017
NOAA HMS	MODIS data	Smoke plumes created by hand by NOAA analysts based on MODIS output	Wettstein et al. 2018; Yao and Henderson 2014
Model-based			
WRF-Chem or TAPM-Chemical transport model	Meteorological conditions and atmospheric chemistry	Inputs meteorological and atmospheric chemistry data to model smoke plumes	Gan et al. 2017 Liu et al. 2017 Haikerwal et al. 2015

All definitions we examined for classifying wildfire smoke days use Washington State $PM_{2.5}$ data from air quality monitors for June-September, 2006-2017. The first two approaches used $PM_{2.5}$ as the exposure metric, with two different thresholds: 15 and 20.4 μg/m^3. Both definitions categorize days with a 24-hr average $PM_{2.5}$ concentration above the threshold as wildfire smoke days, and all days with a 24-hr average $PM_{2.5}$ concentration below the threshold as non-wildfire smoke days. The first definition uses 15 μg/m^3 as the wildfire smoke threshold to align with the previous breakpoint between Good and Moderate air quality, and the previous NAAQS primary standard for $PM_{2.5}$ (US EPA, n.d.). The second definition uses 20.4 μg/m^3 as the wildfire smoke threshold to align with the Washington Air Quality Advisory threshold between Moderate and Unhealthy for Sensitive Groups, again making it a logical threshold from a state public health messaging standpoint (WA DOH; WA ECY, n.d.).

The third definition sets an air monitor specific threshold based on the maximum summertime $PM_{2.5}$ concentrations across a set of years designated as largely not affected by wildfire smoke. Days within these years that were deemed wildfire smoke affected (based on administrative data cataloging) were thrown out, and the highest remaining $PM_{2.5}$ concentration was set as each monitor's wildfire smoke threshold.

The fourth approach defines a 24-hour average $PM_{2.5}$ concentration greater than 20.4 $\mu g/m^3$ as wildfire smoke affected, as background anthropogenic particulate matter across the study period was thought to be below this level. An additional set of criteria was applied to days between 9 and 20.4 $\mu g/m^3$ in order to capture low $PM_{2.5}$ wildfire smoke events. 9 $\mu g/m^3$ was set at the low end because grid cell days below 9 $\mu g/m^3$ are deemed too low to make a meaningful difference with respect to health (U. S. Environmental Protection Agency, 2010), and there is historical evidence of wildfire smoke affected days in rural areas with $PM_{2.5}$ concentrations close to 9 $\mu g/m^3$ in Washington. For the grid cell days between 9 and 20.4 $\mu g/m^3$, the following criteria were applied:

1) The day must be part of an event in which 2 of 3 consecutive days are greater than 9 $\mu g/m^3$;
2) One of the days in the event window must be greater than 15 $\mu g/m^3$;
3) For urban areas (Seattle, Tacoma, Spokane), at least 50% of the air monitors in those areas must be greater than 9 $\mu g/m^3$

These criteria were selected due to the nature of wildfire smoke events in Washington State. The first two criteria were informed by the historical observation that nearly all smoke events span multiple days, and the third criterion was informed by the observation that smoke events tend to affect nearby monitors in a region.

Each of the four definitions attempted to improve upon the previous definitions by reducing false positives and false negatives. Both the 15 μg/m^3 and 20.4 μg/m^3 thresholds treat rural and urban areas identically, when in reality, background anthropogenic $PM_{2.5}$ is much higher in urban areas, introducing substantial misclassification of days. The site-specific approach was based on subjective judgment, and assumed years relatively unaffected by wildfire smoke were representative of other years for all meteorological conditions.

In order to compare classification of wildfire smoke affected days by definition, organic carbon data from the Chemical Speciation (CSN) and Interagency Monitoring of Protected Visual Environments (IMPROVE) Networks in Washington for 2006-2017 was used. Currently, a total of four monitors in Washington collect chemical speciation data every 3 days (1 monitor) or every 6 days (3 monitors). Historically, as many as five speciation monitors operated concurrently and were moved around to different sites in the state over this time period. The dataset used for this analysis contains chemical speciation data from 8 different air monitoring sites (see figure A3 for documentation of when each site provided CSN or IMPROVE data): Cheeka Peak (IMPROVE), Marysville (CSN), Seattle 10th and Weller (CSN), Seattle Beacon Hill (CSN and IMPROVE), Seattle Duwamish (CSN), Tacoma Alexander (CSN), Tacoma L St (CSN), and Yakima (CSN), covering both rural and urban areas. Other IMPROVE sites exist in Washington but were excluded from this analysis because they do not have daily $PM_{2.5}$ measurements. According to our definition of urban areas, the monitors at Seattle 10th and Weller, Seattle Beacon Hill, Seattle Duwamish, Tacoma Alex, and Tacoma L were considered urban, whereas the monitors at Cheeka Peak, Marysville, and Yakima were not considered urban (see Appendix C for full list).

In order to get an initial sense of days that were likely classified incorrectly, we set conservative organic carbon thresholds. The first is meant to serve as a threshold, below which days are likely to be non-wildfire smoke affected, based on measurements from the literature (Hand et al., 2013; Malm et al., 2011; Turpin et al., 2010). The second is meant to serve as a threshold, above which days are likely to be wildfire smoke affected, based on measurements from the literature (Alves et al., 2011; Gan et al., 2017; Jia Coco Liu et al., 2017; Vicente et al., 2013).

To screen for false positives by monitor site, we calculated the percent of days, among days with OC data at the sites listed above, where OC was both below 1.667 $\mu g/m^3$ and categorized as a wildfire smoke day, for each of the four methods. The 1.667 $\mu g/m^3$ threshold, based on the Hand et al. study, reporting 2-3 $\mu g/m^3$ for background particulate organic matter (POM) concentration for the western US, is calculated as 1.8*OC (Hand et al., 2013). Taking a conservative approach, we assumed the upper end of this range (3 $\mu g/m^3$) is equal to the background POM, and found 1.667 $\mu g/m^3$ to be the OC concentration corresponding to the background POM value with no wildfire smoke.

To screen for false negatives, we calculated the percent of days OC was above 5 $\mu g/m^3$ and not categorized as a wildfire smoke day. As stated above, Gan et al. defines a wildfire smoke affected day as exhibiting wildfire smoke-sourced $PM_{2.5} > 10$ $\mu g/m^3$ (Gan et al., 2017). Wildfire smoke-sourced $PM_{2.5}$ was defined as total $PM_{2.5}$ minus background $PM_{2.5}$. For their WRF-Chem based definition, background $PM_{2.5}$ was estimated by the WRF-Chem model, and for their other two definitions, NOAA's Hazard Mapping System was used to identify days without smoke plumes Gan et al. subtracted background $PM_{2.5}$ to yield wildfire smoke-source $PM_{2.5}$ (Gan et al., 2017; Lassman et al., 2017). Based on measurements from the literature, we assume about half of total $PM_{2.5}$ is due to organic carbon (Alves et al., 2011; Jia Coco Liu et al., 2017; Vicente et al., 2013). Using Gan et al.'s threshold of 10 $\mu g/m^3$ for smoke-source $PM_{2.5}$, we set the OC threshold at 5 $\mu g/m^3$. This is likely a conservative estimate, as the 10 $\mu g/m^3$ threshold from Gan et al. is for smoke-affected $PM_{2.5}$ (Gan et al., 2017), whereas the 50% marker for OC as a percentage of $PM_{2.5}$ is for total $PM_{2.5}$ on smoke-affected days, rather than for smoke $PM_{2.5}$ on smoke-affected days (Alves et al., 2011; Jia C Liu & Peng, 2018; Vicente et al., 2013). Both thresholds used in this analysis are based on current measurements in the literature, and are not meant to capture all days improperly misclassified. Rather, their purpose is to serve as indicators for how our definitions performed for days that, based on published OC measurements, should be considered smoke-affected or not.

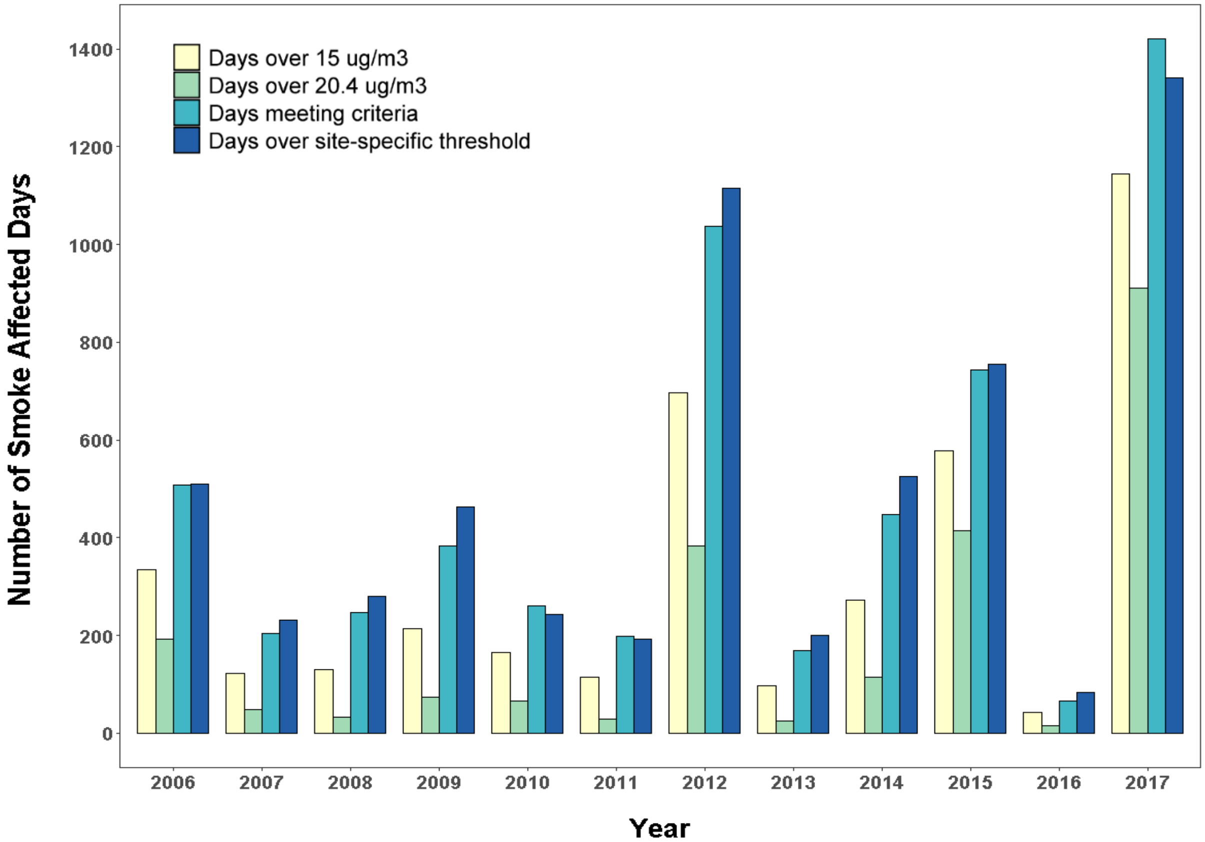
A final calculation of classification agreement was conducted, using the wildfire smoke definitions from Gan et al. as a comparison, as the Gan et al. study is the only existing wildfire smoke and health effects study conducted in Washington State (Gan et al., 2017). The wildfire smoke sourced $PM_{2.5}$ files were downloaded for each of Gan et al.'s three approaches: kriging, geo-weighted ridge regression, and Weather Research and Forecasting with Chemistry (WRF-Chem) chemical-weather model. Classification agreement across the four definitions discussed above compared to Gan et al.'s three definitions was calculated, using only dates that matched in both: all locations in the state of Washington, for July through September, 2012 (Gan et al., 2017). For further discussion of the smoke-sourced $PM_{2.5}$ approaches see Gan et al and their companion paper (Gan et al., 2017; Lassman et al., 2017).

All analyses were performed using R 3.4.3 (R Development Core Team, 2018).

Results

Figure 1 shows the counts of wildfire smoke affected grid cell days (see Appendix C for description of exposure grid cells) by year for each of the four definitions. The plot follows the relative trend we expect: 2012 and 2017, both characterized by statewide wildfire smoke events, have more wildfire smoke affected days than the other years in the time series. Within each year, there is some variability between definitions, suggesting definitions may over- or under-represent wildfire smoke affected days. The 20.4 $\mu g/m^3$ threshold definition (green), yielding the fewest counts of wildfire smoke affected days across all years, likely under counts wildfire smoke days. Across years with fewer wildfire smoke impacts, the site-specific threshold definition (dark blue) and the criteria definition (light blue) yield similar counts. However, in the major wildfire smoke affected years (2012 and 2017) some difference in counts of wildfire smoke affected days is evident.

Figure 1. Number of wildfire smoke affected days by definition



A further comparison of the four definitions is presented in Table 2, showing classification of wildfire smoke and non-wildfire smoke days based on organic carbon thresholds, detailed above (Alves et al., 2011; Gan et al., 2017; Hand et al., 2013; Jia C Liu & Peng, 2018; Vicente et al., 2013). Across all four definitions, 0.1% of days or less both have an organic carbon concentration < 1.667 μg/m^3 and are classified as a wildfire smoke day. However, all definitions yield some days with high organic carbon concentrations that are classified as non-wildfire smoke days. Based on findings from other studies, we assume that all days classified as wildfire smoke affected have an organic carbon concentration > 5 μg/m^3 (Alves et al., 2011; Gan et al., 2017; Jia C Liu & Peng, 2018; Vicente et al., 2013). Across all definitions, 23% or more of days with organic carbon > 5 μg/m^3 are classified as non-wildfire smoke days. This incorrect classification is lowest for the criteria definition, and highest for the 20.4 μg/m^3 threshold definition (Table 2). However, exploratory analysis indicates the 11 days with OC > 5 μg/m^3 classified as non-smoke days for the criteria definition are likely classified correctly. 10 of the 11 days exhibit high elemental carbon (EC) levels (concentration > 0.45 μg/m^3), indicating the elevation of OC is likely due to background sources (Table 2). Figure A1 (appendix A) shows the overlap in distribution of OC for wildfire smoke days versus non-wildfire smoke days by definition, and Figure A2 displays the overlap in distribution of $PM_{2.5}$ for wildfire smoke days versus non-wildfire smoke days by definition.

Table 2. Percent of wildfire smoke and non-wildfire smoke days with OC concentration < 1.667 μg/m^3 and > 5 μg/m^3, by definition

	Days < 1.667 μg/m^3 OC		Days > 5 μg/m^3 OC		Total days misclassified
	Smoke N (%)	**Non-smoke N (%)**	**Smoke N (%)**	**Non-smoke N (%)**	
Criteria definition	0 (0)	778 (100)	36 (76.6)	11 (23.4)	11 (1.3)
Site-specific definition	1 (0.1)	777 (99.9)	34 (72.3)	13 (27.7)	14 (1.7)
15 μg/m^3 threshold definition	1 (0.1)	777 (99.9)	33 (70.2)	14 (29.8)	15 (1.8)
20.4 μg/m^3 threshold definition	0 (0)	778 (100)	17 (36.2)	30 (63.8)	30 (3.6)

A final comparison of the four wildfire smoke definitions to the approaches from Gan et al. (kriging, GWR, and WRF-Chem) is displayed in Table 3 (Gan et al., 2017). Among the days classified as wildfire smoke affected, by definition, 78.3 – 98.2% were also classified as wildfire smoke affected by Gan et al (Table 2; Figure A5). The 20.4 μg/m^3 threshold definition was in most agreement with Gan et al. in classifying wildfire smoke-affected days, while the criteria and the site-specific definition were in most agreement with Gan et al. in classifying non-wildfire smoke affected days (Table 3; Figure A6). Gan et al.'s WRF-Chem model performs worse across all definitions than the kriging and GWR approaches. This is likely due to the difference in source data used to inform the WRF-Chem modeled compared to the kriging and GWR models. While none of the four definitions classifies wildfire smoke and non-wildfire smoke days in complete agreement with Gan et al.'s approaches, the criteria and the site-specific definitions maximize the agreement of both.

Table 3. Percent of wildfire smoke and non-wildfire smoke grid cell days, by definition, classified as wildfire smoke and non-wildfire smoke days by Gan et al., by approach

			Criteria		**Site-specific**		**15 μg/m³ threshold**		**20.4 μg/m³ threshold**	
			WFS	Non WFS	WFS	Non WFS	WFS	Non WFS	WFS	Non WFS
		N	12,428	55,101	13,976	53,553	8,438	59,091	5,002	62,527
			%	%	%	%	%	%	%	%
Kriging	WFS	17,646	89.7	11.8	86.5	10.4	96.6	16.1	97.7	20.4
	Non WFS	49,883	10.3	88.2	13.5	89.6	3.4	83.9	2.3	79.6
GWR	WFS	19,780	91.8	15.2	90.3	13.4	97.4	19.6	98.2	23.8
	Non WFS	47,749	8.2	84.8	9.7	86.6	2.6	80.4	1.8	76.2
WRF-Chem	WFS	40,203	82.4	54.4	78.3	54.6	86.6	55.7	89.3	57.1
	Non WFS	27,326	17.6	45.6	21.7	45.4	13.4	44.3	10.7	42.9

WFS: Wildfire smoke day; Non WFS: Non-wildfire smoke day; GWR: Geo-weighted ridge regression; WRF-Chem: Weather Research and Forecasting with Chemistry

The distribution of $PM_{2.5}$ concentrations was plotted for all differently classified days in Table 3, by definition (Figures A7-10). Figure A7 displays the distribution of $PM_{2.5}$ on days classified as wildfire smoke affected by the criteria definition, that were also classified as non-wildfire smoke affected by each of the three Gan et al. approaches. The figure also shows the reverse: the distribution of $PM_{2.5}$ on days classified as non-wildfire smoke affected by the criteria definition that were also classified as wildfire smoke affected using each of the three Gan et al. approaches. Figures A8-10 show the same distributions for the site-specific, 15 μg/m^3 threshold and 20.4 μg/m^3 threshold, respectively. For days classified as wildfire smoke affected by one of our four definitions, but non-wildfire smoke affected by Gan et al., the site-specific definition shows more days below 10 μg/m^3 classified as wildfire smoke-affected than the criteria definition. On the other hand, the distribution of $PM_{2.5}$ across days classified as non-wildfire smoke affected by one of our four definitions but wildfire smoke affected by Gan et al. are similar across the four definitions.

Appendix Figure A11 shows a final summary of the four definitions compare to Gan et al.'s definitions. Each plot shows the distribution of $PM_{2.5}$ under each of the following scenarios: days where all 7 (our four definitions and Gan et al.'s three methods) definitions classify the day as smoke-affected; days where 6 of 7 definitions classify the day as smoke-affected; days where 5 of 7 definitions classify the day as smoke-affected, and so on, to none of the definitions classifying the day as smoke-affected. We observe fairly separated $PM_{2.5}$ distributions for days categorized as smoke-affected by all 7 definitions compared to by no definitions. The plots showing the distribution of $PM_{2.5}$ on days classified as smoke by only some definitions are what we expect: days that could be considered background $PM_{2.5}$ or wildfire smoke $PM_{2.5}$. This plot further supports use of an approach that sets criteria days in this ambiguous zone must meet for it to be considered wildfire smoke affected, as we have done here.

Discussion

To our knowledge, this is the first study to quantify classification agreement of wildfire smoke exposure definitions. The primary goal of this analysis was an assessment of classification agreement across our four definitions, and against definitions from Gan et al.'s study, as no gold standard for defining a wildfire smoke affected day currently exists. Classification agreement was calculated for the four definitions presented in this analysis using two distinct methods: a comparison of classification using organic carbon speciation data, and classification agreement with definitions from a study in the same location with an overlapping time period (Gan et al., 2017).

The classification comparison method, use of organic carbon speciation data, yielded similar results for false positives across all definitions, but different results for false negatives. We found all four definitions exhibited some days > 5 μg/m^3 OC classified as non-wildfire smoke days. As discussed above, the 11 days in this category under the criteria definition were likely correctly classified as non-wildfire smoke days due to high elemental carbon, indicating other sources of organic carbon. The other three definitions also likely had days in this category, but had additional days that were likely truly misclassified.

Use of OC speciation data as a metric of comparison across wildfire smoke definitions has its limitations. The thresholds we chose have not been used in other studies, but are based on reported measurements and scientific findings. The 1.667 μg/m^3 OC threshold is based on a study reporting background particulate organic matter for the western U.S. (Hand et al., 2013), and the 5 μg/m^3 OC threshold is based on several studies reporting OC constituting about half of total $PM_{2.5}$ (Alves et al., 2011; Jia C Liu & Peng, 2018; Vicente et al., 2013). The 5 μg/m^3 OC threshold is also based on Gan et al.'s threshold for wildfire smoke days of 10 μg/m^3 for wildfire smoke-influenced $PM_{2.5}$. The 5 μg/m^3 OC threshold is likely too low, as the 10 μg/m^3 threshold from Gan et al. is specifically for wildfire smoke-influenced $PM_{2.5}$, and the finding that 50% of $PM_{2.5}$ on wildfire smoke affected days is attributed to OC is based on total $PM_{2.5}$. Thus, the proportion of organic carbon that makes up wildfire smoke-influenced $PM_{2.5}$ is likely greater than 50%. Additionally, the OC speciation data is limited by the number of monitors that collect speciation data in Washington, and the limited frequency of data collection (Figure A3). This results in data that captures a small portion of the state on a subset of the full time series, limiting our ability to accurately compare classification across definitions. Further research is needed to determine if organic carbon measurements and speciation data can be more accurately used to define wildfire smoke affected days through approaches such as positive matrix factorization (Drosatou, Skyllakou, Theodoritsi, & Pandis, 2019; Xie et al., 2012).

The second classification agreement method, a comparison to Gan et al.'s definitions, yields disagreement in classification across all four definitions. The 15 μg/m^3 and 20.4 μg/m^3 threshold definitions exhibit higher agreement of wildfire smoke days than the site specific and criteria definitions across all three approaches from Gan et al., but exhibit much lower agreement of non-wildfire smoke days. In classifying wildfire smoke days, it is important to maximize correct classification of both wildfire smoke affected days and non-wildfire smoke affected days. In the 15 μg/m^3 and 20.4 μg/m^3 methods, 86.6-98.2% of wildfire smoke days are classified as wildfire smoke days by Gan et al.'s definitions. However, only 42.9-83.9% of non-wildfire smoke days are classified as non-wildfire smoke days by Gan et al. Several days we categorized as non-wildfire smoke affected are classified as wildfire smoke affected by Gan et al. The site-specific definition and the criteria definition perform better overall, by minimizing both the percent of wildfire smoke days classified as non-wildfire smoke days by Gan et al. and the percent of non-wildfire smoke days classified as wildfire smoke days by Gan et al.

An accurate calculation of classification agreement across the four definitions, compared to Gan et al.'s approaches, relies on the robustness and accuracy of Gan et al.'s approaches for determining wildfire smoke-influenced $PM_{2.5}$. Gan et al. estimated wildfire smoke-influenced $PM_{2.5}$ using three different methods, yielding similar results across the kriging smoke and GWR smoke models, and different results for the WRF-Chem smoke model (Gan et al., 2017). In the companion paper to the Gan et al. study, Lassman et al. found the WRF-Chem model to have less precision and accuracy in estimating surface level $PM_{2.5}$ compared to the kriging smoke and GWR smoke approaches (Gan et al., 2017; Lassman et al., 2017). Further, Lassman et al. found the WRF-Chem smoke model tended to overestimate $PM_{2.5}$ concentrations compared to the other two models tested, resulting in a smoke exposure incident in July in eastern Washington not captured by the other two definitions (Gan et al., 2017; Lassman et al., 2017). Based on these observations, the WRF-Chem smoke model likely exhibits higher levels of exposure misclassification than the other two approaches, and may be a poor model choice for comparison of smoke day classification. These observations are in line with our results, showing that all four of our definitions have relatively low agreement with the WRF-Chem smoke model in classifying smoke affected days compared to the kriging smoke and GWR smoke models.

We cannot compare Gan et al.'s three smoke classification approaches to a gold standard, and thus cannot determine the exact misclassification of their wildfire smoke exposure definitions. We would expect some differences in classification due to differences in methodology between their approaches, and compared to our definitions, which is most evident in the WRF-Chem smoke model. The large differences in methodology and in exposure classification seen across Gan et al.'s three approaches and our four definitions are precisely the reasons a gold standard for wildfire smoke classification is needed. Different wildfire smoke classification approaches often yield differing effect estimates, and are then directly compared across studies and in review papers when they should not be compared (Gan et al., 2017; Jia C Liu et al., 2015; C E Reid et al., 2016). A standard wildfire smoke classification approach would aid in developing comparable effect estimates across space and time.

Additionally, Gan et al.'s analysis was limited to July-October 2012, limiting our comparison to three months out of 48 total months in the time series. It is possible that this limited time period does not capture the full range and variability in $PM_{2.5}$ seen in the time series, and thus may not accurately capture classification agreement across the full time series. However, no other wildfire smoke health impact studies have been conducted in Washington State, and thus no other directly identical spatiotemporal comparison can be drawn to accurately compare classification approaches.

The criteria definition for classifying wildfire smoke days discussed above adds substantial nuance to the exposure assessment discussion and is relatively straightforward to implement. At the basis of the exposure assessment is $PM_{2.5}$ air quality monitoring data, which are widely and publicly available, inexpensive to collect, and include a long historic record. Other definitions discussed above require highly specialized inputs or speciated data. Those methods, while attractive, are not accessible to a health department who may have interest in classifying days to inform risk messaging and management.

The differences in exposure assessment methods across studies stems largely from differences in classification of wildfire smoke days or time periods. Some studies treat wildfire smoke as a binary event (Dennekamp et al., 2015; F. Johnston et al., 2011; Martin et al., 2013), while others treat wildfire smoke on a continuum (Alman et al., 2016; Hänninen et al., 2009; Hong et al., 2017), including all days that have at least some wildfire smoke influence in their analysis. Further research in wildfire smoke exposure assessment should separate out these approaches, as they answer different questions and their effect estimates should not be directly compared.

More research is needed to quantify classification agreement across other locations and time periods, and to develop a standard for defining wildfire smoke exposure for use across studies for more accurate comparisons of risk estimates for health outcomes of interest (Gan et al., 2017;

Henderson et al., 2011). This study represents the first effort to quantify classification agreement across wildfire smoke day definitions, and provides additional insight into the relative strengths and weaknesses of the approaches explored.

Aim 2: Mortality Associated with Wildfire Smoke Exposure in Washington State, 2006-2017

Mortality Associated with Wildfire Smoke Exposure in Washington State, 2006-2017

Annie Doubleday[1], Jill Schulte[2], Ranil Dhammapala[2], Lianne Sheppard[1,3], Matt Kadlec[2], Julie Fox[4], Tania Busch Isaksen[1]

[1]Department of Environmental and Occupational Health Sciences, University of Washington
[2]Air Quality Program, Washington State Department of Ecology
[3]Department of Biostatistics, University of Washington
[4]Environmental Public Health Sciences, Washington State Department of Health

Abstract

Background: Wildfire events are increasing in prevalence in the western United States. Research has found mixed results on the degree to which exposure to wildfire smoke increases the risk of mortality.

Objectives: We tested for an association between exposure to $PM_{2.5}$ on wildfire smoke days and increased odds of all non-traumatic causes of mortality in Washington State.

Methods: We characterized wildfire smoke exposure by daily average $PM_{2.5}$ concentrations on wildfire smoke days, from June 1 through September 30, for 2006-2017. Wildfire smoke days were defined as all days with nearest monitor ambient concentration above a $PM_{2.5}$ value of 20.4 $\mu g/m^3$, with an additional set of criteria applied to days between 9 and 20.4 $\mu g/m^3$. We employed a case crossover study design using conditional logistic regression and time-stratified referent sampling to estimate the $PM_{2.5}$ association with mortality. Monitoring data was obtained from the Washington State Department of Ecology's monitoring network, and mortality data was obtained from the Washington State Department of Health.

Results: After adjusting for temperature and humidity, the odds of same-day non-traumatic mortality were 1.0% greater on wildfire smoke days compared to non-wildfire smoke days (95%

CI: 0.99 – 1.04). Among adults ages 65-84, we estimated a 2.0% increase in the odds of same-day non-traumatic mortality (95% CI: 0.98 – 1.06), and an 8.0% increase in the odds of same-day respiratory mortality on wildfire smoke days compared to non-wildfire smoke days in this age group (95% CI: 0.96 – 1.21). Among all ages, we estimated a 1.3% increase in the odds of non-traumatic mortality due to previous day wildfire smoke exposure compared to no wildfire smoke exposure (95% CI: 1.00 – 1.02).

Discussion: This study is the first to examine wildfire smoke and mortality in Washington State, and is consistent with other research on wildfire smoke and mortality. This study lays the groundwork for estimating the risk associated with exposure to wildfire smoke in Washington, and will help inform local and state risk communication efforts and decision-making during future wildfire smoke events.

Introduction

The western United States is seeing an increase in wildfires during the summer and fall months, increasing the importance of understanding the health impacts associated with exposure to wildfire smoke (Dennison et al., 2014; Westerling, 2016). It is estimated that the total forest fire area burned nearly doubled during 1984-2015 compared to the area projected to have burned without climate change (Abatzoglou & Williams, 2016). This trend is expected to worsen, with climate projections indicating wildfires in the western U.S. will increase in frequency and intensity (Joyce et al., 2014; Littell et al., 2009). The Intergovernmental Panel on Climate Change (IPCC) estimates that climate change will increase the wildfire season in North America by 10-30% (Parry, 2007). This is expected to result in worsening air quality during wildfire season in the coming decades (Interagency Working Group on Climate Change and Health, 2010).

Wildfire smoke contains a wide range of compounds known to be harmful to human health, including fine particulate matter ($PM_{2.5}$), acrolein, benzene, carbon monoxide, and polycyclic aromatic hydrocarbons (Naeher et al., 2010; Urbanski et al., 2008). Exposure to these toxic compounds is of concern near the source, as well as several hundred miles away (Moeltner et al., 2013; Munoz-Alpizar et al., 2016). Ash from wildfires can be carried hundreds of miles by water, contaminating soil and water sources (Hohner, Cawley, Oropeza, Summers, & Rosario-ortiz, 2016). While it has been shown that the toxic compounds from wildfire smoke travel long distances from the source, potentially exposing thousands of individuals, the health effects associated with exposure are just beginning to be understood (Black et al., 2017).

Exposure assessment is challenging, as there is no agreed upon gold standard method for defining what constitutes a wildfire smoke affected day used in the existing health effects literature. Common methods utilize air quality area monitoring particulate matter (PM) measurements, satellite data, air pollution chemistry modeling, or a combination of methods to create a more robust definition. The first and most commonly used method to assess exposure to wildfire smoke is through PM monitoring data, typically PM less than 2.5 microns in diameter ($PM_{2.5}$) or less than 10 microns (PM_{10}) (Alman et al., 2016; Dennekamp et al., 2015; Haikerwal et al., 2015; Hänninen et al., 2009; Hong et al., 2017; F. Johnston et al., 2011; Linares et al., 2018, 2015; Sastry, 2002; Shaposhnikov et al., 2014; Zu et al., 2016). Within this method, some studies examine the health impact for each $10\mu g/m^3$ increase in $PM_{2.5}$ (Alman et al., 2016; Hänninen et al., 2009), while other studies examine the impact above a specific threshold, such as the 99th percentile of $PM_{2.5}$ for the time series (F. Johnston et al., 2011; Martin et al., 2013). The second method of determining wildfire smoke exposure is to use Moderate Resolution Imaging Spectroradiometer (MODIS) satellite imagery data, such as Aerosol Optical Depth (AOD) or Fire Radiative Power (FRP) to serve as a proxy for PM or to determine smoke days based on the distribution of AOD or FRP over the time series (Delfino et al., 2009; Elliott et al., 2013; Faustini et al., 2015; Yao et al., 2016). Additionally, MODIS imagery is used to generate the NOAA Hazard Management System (HMS) fire and smoke product, to serve as a binary indicator of smoke presence on a daily basis (National Oceanic and Atmospheric Administration, 2018; Wettstein et al., 2018; Yao & Henderson, 2014). These methods are often used when reliable and adequate air monitoring measurements are unobtainable. The final method commonly used is to simulate daily $PM_{2.5}$ concentrations using the Weather Research and Forecasting with Chemistry (WRF-Chem) model (Grell et al., 2005). The WRF-Chem model combines meteorological conditions with atmospheric chemical processes to simulate air quality, and daily $PM_{2.5}$ concentrations. Several studies combine the above methods to increase the accuracy of their wildfire smoke day definition, or to compare to their primary method of assessment (Dell et al., 2019; Henderson et al., 2011; Kollanus et al., 2016; Martin et al., 2013; Yao & Henderson, 2014). However, because no gold standard method exists, it is not clear that combining methods necessarily yields more accurate estimation of wildfire smoke attributable days. Further, differences in effect estimates across studies may, in part, be due to differences in exposure assessment approach. More research in determining wildfire smoke affected days needs to be conducted to validate methods currently in use.

Using the exposure assessment methods outlined above, evidence for an association between exposure to wildfire smoke and all-age mortality is mixed. A handful of studies provide evidence for an association between short-term exposure to wildfire smoke and increased risk of mortality for all ages (Haikerwal et al., 2015; Hänninen et al., 2009; F. Johnston et al., 2011; Linares et al., 2015; Vila et al., 2013), while others do not (Faustini et al., 2015; Hong et al., 2017; F. Johnston et al., 2011; Kollanus et al., 2017). One study in Finland reported a 1.0% (95% CI: 0.8 – 2.1%) increase in all-cause mortality per 10μg/m^3 increase in same-day $PM_{2.5}$ concentration (Hänninen et al., 2009), while another study in Australia estimated that wildfire events were associated with a 2.0% (95% CI: -2.0 – 5.0%) increase in the odds of same-day non-accidental mortality (F. Johnston et al., 2011). Several studies also report estimates for respiratory and cardiovascular mortality, reporting no evidence of association with respiratory mortality (Faustini et al., 2015; F. Johnston et al., 2011; Morgan et al., 2010) and small increases in odds of cardiovascular mortality (Faustini et al., 2015; F. Johnston et al., 2011).

A subset of these studies has reported significant associations between wildfire smoke exposure and non-traumatic mortality in groups 65 years and older, but less of an effect in other age groups (Analitis et al., 2012; Haikerwal et al., 2015; Shaposhnikov et al., 2014; Vila et al., 2013). Analitis et al. (2011) reports that the effect of respiratory mortality in Greece was greater in adults 75 and over for a large fire (Analitis et al., 2012). Further, Haikerwal et al. (2015) observed an increase in risk of cardiac arrests, especially in older adults in Australia, and Vila et al. (2013) reports that older adults in their study in Brazil have the strongest association between exposure to biomass burning and circulatory disease mortality (Haikerwal et al., 2015; Vila et al., 2013). There are few US-based mortality studies, providing little evidence for US-specific mortality associated with wildfire smoke exposure. Further, there are no mortality studies in the Pacific Northwest, necessitating further research on the association between wildfire smoke exposure and non-traumatic mortality in Washington State.

Only two efforts have examined health effects of wildfire smoke in Washington State, both focused on the extreme fire season of 2012 and with limited geographical coverage to Central Washington (Gan et al., 2017; Washington State Department of Health, 2012). A group of local health departments in Central Washington in conjunction with the state Department of Health examined cardiopulmonary emergency department (ED) and outpatient visits in the affected areas before and after a major wildfire smoke event, and found a substantial increase after the wildfire smoke event compared to before. They found a 28% (95% CI: 18 – 39%) increase for cardiovascular disease and respiratory outpatient visits in the two weeks during the wildfires compared to the two weeks before, and an 18% (95% CI: 9 – 28%) increase in the two weeks after the wildfires compared to the two weeks before. While providing evidence for health effects from wildfire smoke exposure in Washington, this study was limited by its small sample size, and did not examine the possible effect on mortality (Washington State Department of Health, 2012). Gan et al. conducted a time-stratified case-crossover study of the same event, and found a 7.6% (95% CI: 1.9 – 13.6%) increase in the risk of cardiopulmonary hospitalizations for each 10 μg/m^3 increase in wildfire smoke $PM_{2.5}$ using exposure based on geographically weighted ridge regression that includes input from the WRF-Chem model, satellite observed aerosol optical depth and kriged $PM_{2.5}$ (Gan et al., 2017). While both of these studies offer insight into the health effects of wildfire smoke exposure in Washington, they are limited in duration and sample size, and only examine hospitalizations. No studies have been conducted on the potential risk of mortality associated with wildfire smoke exposure in Washington, with the closest mortality study conducted in British Columbia in 2017 (Hong et al., 2017). Our study examines the association between wildfire smoke exposure and the odds of non-traumatic mortality in Washington State over 12 years using an exposure assessment method that reduces misclassification. We hypothesized *a priori* that evidence of an elevated risk would be limited to individuals 65 and over, to an effect at a lag of one or more days, and within respiratory and cardiovascular causes, consistent with many of the effects seen in the literature discussed above.

Methods

<u>Health outcome data</u>

This study was conducted using Washington State mortality and exposure data from June through September, 2006 to 2017. We obtained mortality data from the Washington State Department of Health for years 2006-2017. Deaths in June through September were included to focus on mortality during Washington wildfire season only. Washington mortality data includes latitude and longitude of death, underlying cause of death, date of death, and other individual characteristics. We examined non-traumatic causes of death only (ICD-10 codes: A01-V99), and excluded traumatic causes of death likely to be unrelated to wildfire smoke exposure. Within non-traumatic causes of death, we examined cardiovascular causes (ICD-10 codes: I05-I52), including ischemic heart disease (ICD-10 codes: I20-22, 24), respiratory causes (ICD-10 codes: J01-J99), including asthma (ICD-10 codes: J45-46), COPD (ICD-10 codes: J41-44), pneumonia (ICD-10 codes: J12-18), and cerebrovascular causes, which includes ischemic stroke (ICD-10 codes: I60-67). About 97% of cases in the mortality data include the latitude and longitude of death; we excluded the remaining 3% from the analysis. We used the geo-coded mortality data to attribute each death to an exposure grid cell in order to assign $PM_{2.5}$ exposure to each case. The main outcome of interest was non-traumatic mortality on wildfire smoke days compared to non-wildfire smoke days, and stratified by specific causes of death.

Other individual characteristics (race, education, age, sex) were included in the mortality data files. Median household income at the census tract level was obtained from the U.S. Census Bureau for 2010-2017. Median household income was extrapolated to 2006, based on the five-year average percent change in growth within each census tract. For the median household income analysis only, we omitted 10% of cases due to missing census tract or missing median household income data.

Exposure data

This study derived its exposure data using empirical data collected from regulatory area air quality monitors combined with the AIRPACT (Air Indicator Report for Public Awareness and Community Tracking) model to produce $PM_{2.5}$ gridded exposure data. Air quality in Washington State is monitored with 68 federal and state-run monitors that collect 24-hour air quality metrics, including $PM_{2.5}$, NO_2, and CO, and meteorological variables, including temperature, dew point, and wind speed (Ragan, 2018). The 4x4km exposure grid is based on grid cells used in the AIRPACT model, which models air quality at high resolution across the

Pacific Northwest (Washington State University, 2019). We used daily average temperature and dew point to calculate humidex, a measure of apparent temperature calculated from air temperature and dew point (Masterton & Richardson, 1979), for each monitor. Each air quality monitoring site was matched to the AIRPACT model grid cell closest to it. Each monitoring site was then assigned a ratio from the each AIRPACT grid yearly summertime $PM_{2.5}$ mean and each monitor summertime 24-hour $PM_{2.5}$ mean. These ratios, along with the calculated humidex values, were interpolated across the state using Empirical Bayesian Kriging (Pilz & Spöck, 2008). The interpolated ratio at each 4x4km grid cell was then multiplied by the modeled 2014-2017 summertime mean $PM_{2.5}$ to end up with the estimated summertime means across the state, at a 4x4km resolution. The 4x4km $PM_{2.5}$ grid was then overlaid with the Washington State Office of Financial Management's (OFM) yearly population estimates at the census block group level (Management, 2017). In cases where census block group boundaries cross grid cells, we determined the percent of the population's census block group attributed to each grid cell by the area that fell within the grid cell. This method assumes populations are evenly distributed within each census block group, and the population within each grid cell lives at the centroid of the grid cell. Grid cells were assigned the $PM_{2.5}$ concentration of the closest monitor on each day, unless the difference between the summertime mean of the grid cell (determined above) and the nearest monitor exceeded 2 $\mu g/m^3$, or if the monitor was inactive on that day. In those cases, the grid cell was assigned the next closest eligible monitor. The end result is a dataset with the following for each day and each grid cell: 24-hour average $PM_{2.5}$ concentration and humidex from a neighboring monitoring site with the population attributed to that grid cell. This dataset was then joined with the above described health outcome data using a spatial join in ArcGIS (version

10.5.1) (Esri, Redlands, CA), assigning each latitude and longitude of death to a grid cell and corresponding $PM_{2.5}$ concentration and humidex, in degrees Celsius.

<u>Wildfire Smoke day classification</u>

In order to classify wildfire smoke affected days, a number of approaches were considered to minimize false positives in urban areas that experience high background particulate matter concentrations, and to minimize false negatives in rural areas that experience low background particulate matter concentrations (Dockery et al., 1993; U. S. Environmental Protection Agency, 2010). Wildfire smoke affected days are defined as grid cell days with a 24-hour average $PM_{2.5}$ concentration greater than 20.4 μg/m^3, as background anthropogenic particulate matter across the study period was thought to be below this level. An additional set of criteria was applied to days between 9 and 20.4 μg/m^3 in order to capture low $PM_{2.5}$ wildfire smoke events. 9 μg/m^3 was set at the low end because grid cell days below 9 μg/m^3 are deemed too low to make a significant difference with respect to health (U. S. Environmental Protection Agency, 2010), and there is historical evidence of wildfire smoke affected days in rural areas with $PM_{2.5}$ concentrations close to 9 μg/m^3 in Washington. For the grid cell days between 9 and 20.4 μg/m^3, the following criteria were applied:

1) The day must be part of an event in which 2 of 3 consecutive days are greater than 9 μg/m^3;
2) One of the days in the event window must be greater than 15 μg/m^3;
3) For urban areas (Seattle, Tacoma, Spokane), at least 50% of the air monitors in those areas must be greater than 9 μg/m^3

This definition minimized false positives in urban areas while recouping some of the false negatives in rural areas, increasing the accuracy of the wildfire smoke affected day definition.

Statistical analysis

A time-stratified case crossover design was employed to examine the association between wildfire smoke exposure and non-traumatic mortality with conditional logistic regression models. This study design compares wildfire smoke exposure on the day of death to wildfire smoke exposure on several non-event days (referent, or control days) for the same individual. Referent days are selected using time-stratified sampling, matching by the same day of the week, month, and year of death, yielding 3.39 referent days per individual, on average (Table 2). By design, this controls for time-invariant confounders, including sex, age, race, pre-existing health conditions, and other individual characteristics and risk factors, as individuals serve as their own control (Lumley & Levy, 2000). This design controls for some time-dependent variables based on the referent selection method, including day of the week, and seasonal trends in air pollution (Janes, Sheppard, & Lumley, 2005). Results are reported as a percent change in odds of non-traumatic mortality for wildfire smoke-affected days versus non-wildfire smoke affected days, after controlling for humidex. We adjusted for humidex by adding a term into the conditional logistic regression estimating equation for all analyses. The *clogit* function in the *survival* R package was used to conduct the regression analysis (R Development Core Team, 2018).

To examine the effect of wildfire smoke exposure by characteristics of interest, additional subgroup analyses were conducted, stratifying by age group, race category, cause of death, and location (urban versus rural). A lag analysis was conducted using an unconstrained distributed lag model from days 0-4, with day 0 modeled as the wildfire smoke affected day, lag day 1 as the previous day, and so on.

Additional secondary lag analyses were conducted, first for same-day exposures, and second for previous day exposures. For same-day exposures, estimates were reported by age group and race within respiratory causes of death, and by age group for COPD causes of death. For previous day exposures, estimates were reported by cause of death (Table 6). A sensitivity analysis was conducted, setting 20.4 $\mu g/m^3$ as the wildfire smoke threshold to assess whether our results were sensitive to the exposure definition (Table B1).

All analyses were performed using R 3.4.3 (R Foundation, Vienna, Austria) (R Development Core Team, 2018). The University of Washington Institutional Review Board reviewed the study design and determined the study to be exempt. The Washington State Institutional Review Board determined that we did not need to submit a form for official determination of exempt status for use of Washington State mortality data for this study.

Results

Characteristics of the 171,804 non-traumatic deaths in Washington from June through September, 2006-2017 are shown in Table 4. Most non-traumatic deaths occurred in individuals 65 years and older (76.7%), most were white (90.3%), most had a high school degree or less (56.3%), and most lived in census tracts with a median household income of less than $75,000 (70.4%). About a quarter of deaths were due to cardiovascular causes (25.9%), less than 10% were due to respiratory causes (9.5%), and the remaining were due to other non-traumatic causes. There was no trend by day of the week, with deaths occurring fairly evenly across days. About half of all non-traumatic deaths during this period occurred in urban areas (Seattle, Tacoma, and Spokane) (48.9%), with the remaining deaths occurring in all other locations across the state. Table 4 reports the number and percent of strata that contribute to the inferential analysis, defined as those exhibiting exposure contrast with both wildfire smoke and non-wildfire smoke days in their strata.

Table 4. Non-traumatic mortality characteristics

Characteristic		N (%)	N (%) with exposure contrast[1]
Total		**171,804 (100)**	**31,719**
Age group (years)			
	0-4	2,301 (1.3)	423 (1.3)
	5-14	668 (0.4)	140 (0.4)
	15-44	4,969 (2.9)	935 (2.9)
	45-64	32,375 (18.7)	6,082 (19.2)
	65-84	75,110 (43.4)	13,723 (43.3)
	85+	57,618 (33.3)	10,416 (32.8)
Death day of week			
	Monday	24,476 (14.1)	4,626 (14.6)
	Tuesday	24,517 (14.2)	5,047 (15.9)
	Wednesday	24,592 (14.2)	4,934 (15.6)
	Thursday	24,800 (14.3)	4,690 (14.8)
	Friday	25,173 (14.5)	3,865 (12.2)
	Saturday	25,070 (14.5)	4,401 (13.9)
	Sunday	24,413 (14.1)	4,156 (13.1)
Location			
	Non-urban	88,415 (51.1)	17,770 (56.0)
	Urban	84,626 (48.9)	13,949 (44.0)
Median household income by census tract			
	<$35,000	16,092 (9.4)	3,668 (11.6)
	$35,000 - $50,000	41,919 (24.4)	7,921 (25.0)
	$50,000 - $75,000	63,056 (36.7)	11,037 (34.8)
	$75,000 - $100,000	24,631 (14.3)	4,579 (14.4)
	≥$100,000	8,171 (4.8)	1443 (4.5)
	Not reported	17,935 (10.4)	3,071 (9.7)
Race			
	Asian	5,215 (3.0)	943 (3.0)
	Black	4,686 (2.7)	885 (2.8)
	Hispanic	2,230 (1.3)	580 (1.8)
	Native American	2,449 (1.4)	472 (1.5)
	Native Hawaiian/ Other Pacific Islander	1,635 (0.9)	329 (1.0)
	White	156,267 (90.3)	28,395 (89.5)

	Not reported	559 (0.3)	115 (0.4)
Sex			
	Female	86,874 (50.6)	15,946 (50.3)
	Male	84,928 (49.4)	15,772 (49.7)
	Not reported	2 (0)	1 (0)
Underlying cause of death			
	Cardiovascular	44,565 (25.9)	8,135 (25.6)
	Ischemic heart disease	7,945 (4.6)	1,482 (4.7)
	Respiratory	16,286 (9.5)	2945 (9.3)
	Asthma	254 (0.1)	46 (0.1)
	COPD	9,571 (5.6)	1,732 (5.5)
	Pneumonia	2,174 (1.3)	380 (1.2)
	Cerebrovascular[2]	3,747 (2.2)	710 (2.2)

[1]Percent of cases whose strata have both wildfire smoke days and non-wildfire smoke days
[2]Cerebrovascular causes of death include ischemic stroke

Table 5 displays the daily and mean exposure characteristics for $PM_{2.5}$ across the study period. Washington is split into 10,106 4x4km grid cells, represented by 67 air quality monitors. On average, each case had 3.39 referent days (SD, 0.53). The average $PM_{2.5}$ concentration on event days (days of death) was 6.38 μg/m^3 (SD, 9.28), and the average $PM_{2.5}$ concentration on referent days was 6.35 μg/m^3 (SD, 9.11) (Figure B1). The average $PM_{2.5}$ concentration on wildfire smoke days was 26.4 μg/m^3 (SD, 31.9), and on non-wildfire smoke days was 4.67 μg/m^3 (SD, 2.53) (Figure B2). The average humidex on wildfire smoke days was 29.9 (SD, 5.53) in degrees C, and on non-wildfire smoke days was 24.9 (SD, 6.03) in degrees C (Figure B3). Over the 12-year time series, there were 1,031,804 wildfire smoke affected grid cell days, representing about 7.0% of all grid cell days in the time series.

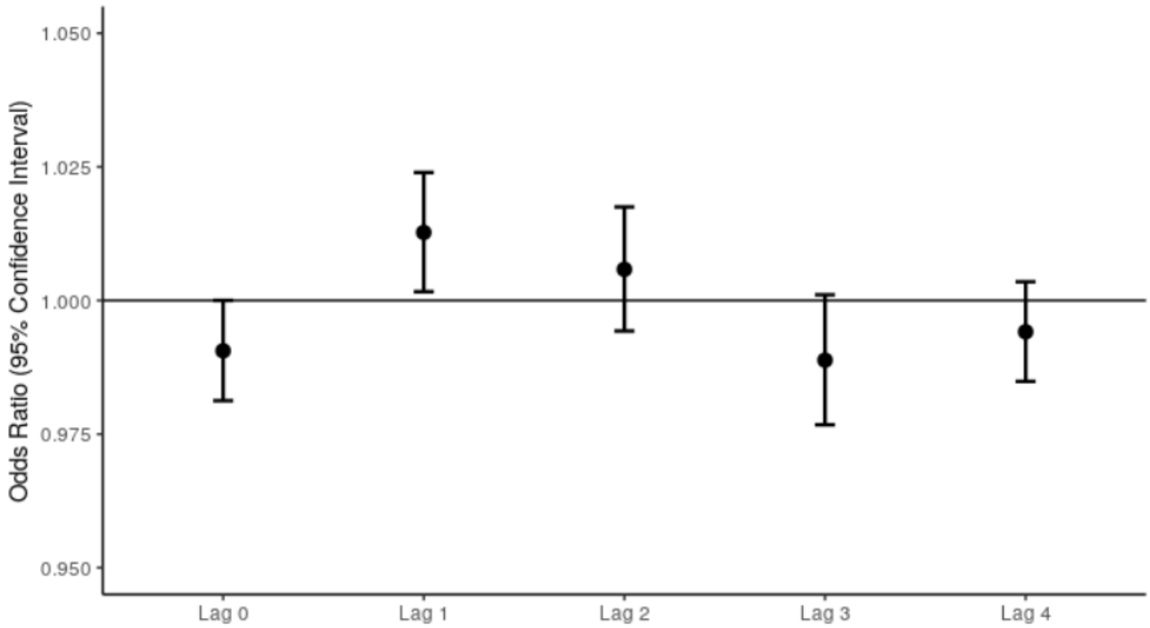
Table 5. Daily $PM_{2.5}$ characteristics for mortality days and referent days

Characteristic	Number
Number of exposure grid cells	10,106
Number of wildfire smoke grid cell days	1,031,804
Total event days	170,854
Total referent days	582,609

Characteristic		$PM_{2.5}$ ($\mu g/m^3$) Mean (SD)
Event days		6.38 (9.28)
Referent days		6.35 (9.11)
Mean referent days per case		3.39 (0.53)
Exposure metric	**$PM_{2.5}$ ($\mu g/m^3$) Mean (SD)**	**Humidex (°C) Mean (SD)**
Wildfire Smoke Days	26.4 (31.9)	29.9 (5.53)
Non-Wildfire Smoke Days	4.67 (2.53)	24.9 (6.03)

Table 6 presents results of the inferential analysis, indicating a 1.0% (95% CI: 0.99 – 1.04) increase in the odds of all same-day non-traumatic mortality on wildfire smoke days versus non-wildfire smoke days, controlling for humidex. We further observed a 1.0% (95% CI: 0.94 – 1.04) decrease in the odds of same-day cardiovascular mortality and a 9.0% (95% CI: 1.00 – 1.18) increase in the odds of same-day respiratory mortality on wildfire smoke days compared to non-wildfire smoke days, controlling for humidex. Among ages 65-84, we observed a 2.0% (95% CI: 0.98 – 1.06) increase in the odds of all same-day non-traumatic mortality on wildfire smoke days versus non-wildfire smoke days, controlling for humidex. The results of an unconstrained distributed lag analysis examine the effect of exposure to wildfire smoke in the four days prior to death. We estimate a 1.3% (95% CI: 1.00, 1.02) increase in the odds of next-day non-traumatic mortality on wildfire smoke days versus non-wildfire smoke days, controlling for humidex and exposure on the four days prior to death (Figure 2). Figure 2 further shows evidence for a lagged effect on the two days prior to death, taking into account the effect of all other lag days. The results do not provide evidence for a lagged effect prior to two days before death.

Figure 2. ORs from an unconstrained distributed lag model by lag day



Additional same-day analyses were conducted, stratifying by age group, location, census tract median household income, race, sex, and cause of death (Table 6). The majority of these analyses indicate little evidence for an effect. Several subgroups report narrow confidence intervals, albeit with odds ratios near 1.0, indicating no change in the odds of mortality by subgroup on wildfire smoke days compared to non-wildfire smoke days. However, within the stratification by race category, we observe a 19.0% (95% CI: 0.67 – 0.99) decrease in the odds of same-day non-traumatic mortality among Hispanics on wildfire smoke days compared to non-wildfire smoke days, controlling for humidex. When stratifying by cause of death, we observe a 9.0% (95% CI: 1.00 – 1.18) increase in the odds of same-day respiratory mortality and a 14.0% (95% CI: 1.02 – 1.26) increase in the odds of same-day COPD mortality on wildfire smoke days compared to non-wildfire smoke days, controlling for humidex.

Table 6. Odds ratios (ORs) and 95% confidence intervals (CIs) of non-traumatic mortality

Category		Adjusted OR (95% CI)	N (%)
All non-traumatic		1.01 (0.99, 1.04)	171,804 (100)
Age group (years)			
	0-4	0.97 (0.77, 1.22)	2,301 (1.3)
	5-14	0.95 (0.64, 1.41)	668 (0.4)
	15-44	0.99 (0.85, 1.15)	4,969 (2.9)
	45-64	1.00 (0.95, 1.06)	32,375 (18.7)
	65-84	1.02 (0.98, 1.06)	75,110 (43.4)
	85+	1.00 (0.96, 1.05)	57,618 (33.3)
Location			
	Non-urban	1.00 (0.96, 1.03)	88,415 (51.1)
	Urban	1.02 (0.99,1.06)	84,626 (48.9)
Median household income by census tract†			
	<$35,000	0.99 (0.92, 1.07)	16,092 (9.4)
	$35,000 - $50,000	1.05 (0.99, 1.10)	41,919 (24.4)
	$50,000 - $75,000	1.00 (0.96, 1.04)	63,056 (36.7)
	$75,000 - $100,000	1.03 (0.96, 1.10)	24,631 (14.3)
	≥$100,000	0.98 (0.87, 1,11)	8,171 (4.8)
Race			
	White	1.01 (0.99, 1.04)	156,267 (90.3)
	Black	1.04 (0.89, 1.21)	4,686 (2.7)
	Native American	0.95 (0.77, 1.17)	2,449 (1.4)
	Hispanic	0.81 (0.67, 0.99)*	2,230 (1.3)
	Native Hawaiian/ Other Pacific Islander	1.19 (0.93, 1.52)	1,635 (0.9)
	Asian	0.98 (0.84, 1.14)	5,215 (3.0)
Sex			
	Female	1.00 (0.96, 1.03)	86,874 (50.6)
	Male	1.03 (0.99, 1.06)	84,928 (49.4)
Underlying cause of death			
	Cardiovascular	0.99 (0.94, 1.04)	44,565 (25.9)
	Ischemic heart disease	1.04 (0.93, 1.17)	7,945 (4.6)
	Respiratory	1.09 (1.00, 1.18)	16,286 (9.5)
	Asthma	0.51 (0.23, 1.12)	254 (0.14)
	COPD	1.14 (1.02, 1.26)*	9,571 (5.6)
	Pneumonia	1.08 (0.86, 1.36)	2,174 (1.3)
	Cerebrovascular	0.90 (0.75, 1.07)	3,747 (2.2)

* $p \leq 0.05$, ** $p \leq 0.01$, *** $p \leq 0.001$; †Annual median household income estimates at the census tract level

Secondary analyses were conducted for age group and race within respiratory causes of death with same-day exposure, and separately for cause of death with the previous day's exposure. For ages 45-64, we observed a 35.0% (95% CI: 1.09 – 1.67) increase in the odds of respiratory mortality, and 33.0% (95% CI: 1.00 – 1.78) increase in the odds of COPD mortality on wildfire smoke days compared to non-wildfire smoke days, controlling for humidex. Within all respiratory mortality, we observed a 12.0% (95% CI: 1.02 – 1.22) increase in the odds of same-day respiratory mortality among whites on wildfire smoke days compared to non-wildfire smoke days, controlling for humidex. At a lag of one day, we observed a 2.0% (95% CI, 1.00 – 1.05) increase in the odds of all non-traumatic mortality, and a 5.0% (95% CI: 0.97 – 1.15) increase in the odds of respiratory mortality on wildfire smoke days, compared to non-wildfire smoke days. Table 7 displays the results from the extended set of secondary analyses.

Table 7. Odds ratios (ORs) and 95% confidence intervals (CIs) for secondary analyses

Category	Adjusted OR (95% CI)	N (%)
Day 0 Results		
Respiratory causes of death, lag day 0		
Age group, for respiratory causes of death		
0-4	1.52 (0.58, 3.97)	119 (0.7)
5-14	-	28 (0.2)
15-44	0.91 (0.45, 1.84)	217 (1.3)
45-64	1.35 (1.09, 1.67)*	2,152 (13.2)
65-84	1.08 (0.96, 1.21)	8,489 (52.1)
85+	1.00 (0.86, 1.16)	5,281 (32.4)
Race, for respiratory causes of death		
White	1.12 (1.02, 1.22)*	15,104 (92.7)
Black	0.96 (0.53, 1.72)	301 (1.9)
Native American	0.88 (0.43, 1.82)	227 (1.4)
Hispanic	0.42 (0.19, 0.95)*	146 (0.9)
Native Hawaiian/ Other Pacific Islander	0.70 (0.20, 2.41)	107 (0.7)
Asian	0.88 (0.52, 1.53)	351 (2.2)
Age group, for COPD causes of death		
45-64	1.33 (1.00, 1.78)	1,281 (13.4)
65-84	1.14 (0.99, 1.31)	5,654 (59.1)

	85+	1.04 (0.85, 1.28)	2,584 (27.0)
Lag day 1 results			
All non-traumatic, lag day 1		1.02 (1.00, 1.05)	171,804 (100)
Cause of death, lag day 1			
	Cardiovascular	1.02 (0.97, 1.07)	44,565 (25.9)
	Ischemic heart disease	1.00 (0.89, 1.13)	7,945 (4.6)
	Respiratory	1.05 (0.97, 1.15)	16,286 (9.5)
	Asthma	0.65 (0.31, 1.35)	254 (0.14)
	COPD	1.07 (0.96, 1.20)	9,571 (5.6)
	Pneumonia	0.97 (0.77, 1.22)	2,174 (1.3)
	Cerebrovascular	0.88 (0.74, 1.05)	3,747 (2.2)

* $p \leq 0.05$, ** $p \leq 0.01$, *** $p \leq 0.001$

Discussion

This is the first study examining the association between wildfire smoke exposure and mortality in Washington State, where little research on other health effects of wildfire smoke has been conducted (Gan et al., 2017; Washington State Department of Health, 2012). We found a 1.0% (95% CI: 0.99 – 1.04) increase in the odds of non-traumatic mortality with same-day wildfire smoke exposure. Other studies examining the association between wildfire smoke exposure and non-traumatic mortality report similar results. One study in southern Europe reports a 1.78% (95% CI: -0.91 – 4.53%) increase in natural mortality (lag 0-1) on smoky days compared to non-smoky days (Faustini et al., 2015), and a study in Finland reports a 0.2% (95% CI: -7.6 – 8.6%) increase in non-accidental mortality for a 10 $\mu g/m^3$ increase in $PM_{2.5}$ on smoke days (Kollanus et al., 2016), and a third study in Australia reports a 2.0% (95% CI: 0.98 – 1.05) increase in the odds of same-day non-accidental mortality (F. Johnston et al., 2011). A final study in British Columbia reports a 1.2% (95% CI: -0.3 – 2.6%) increase in non-accidental mortality associated with daily $PM_{2.5}$ during wildfire season (Hong et al., 2017). As with our results, these studies found small increases in same-day non-traumatic or non-accidental mortality on wildfire smoke days compared to non-wildfire smoke days.

In additional primary analyses, we reported a 1.0% (95% CI: 0.94 – 1.04) decrease in the odds of same-day cardiovascular mortality and a 9.0% (95% CI: 1.00 – 1.18) increase in the odds of same-day respiratory mortality on wildfire smoke days compared to non-wildfire smoke days, controlling for humidex. Some studies examining same-day cardiovascular and respiratory mortality report similar results (F. Johnston et al., 2011; Morgan et al., 2010), while others do not (Faustini et al., 2015; F. Johnston et al., 2011; Morgan et al., 2010). Johnston et al. report a 6.0% (95% CI: 0.97 – 1.17) increase in the odds of same-day cardiovascular mortality, and no (95% CI: 0.79 – 1.25) increase in the odds of same-day respiratory mortality on wildfire smoke days versus non-wildfire smoke days (F. Johnston et al., 2011). Morgan et al. reported almost no change in cardiovascular or respiratory mortality, with a 0.76% (95% CI: -0.76 – 2.30%) increase in same-day cardiovascular mortality, and a 0.32% (95% CI: -3.70 – 3.18%) decrease in same-day respiratory mortality for a 10 $\mu g/m^3$ increase in bushfire PM_{10} (Morgan et al., 2010). Finally, Faustini et al. reported a 6.29% (95% CI: 1.00 – 11.85%) increase in cardiovascular mortality in the five days prior to death, and a -3.49% (95% CI: -9.60 – 3.03%) decrease in respiratory mortality in the five days prior to death for smoke-affected days compared to non-smoke affected days, pooled across 10 cities (Faustini et al., 2015). However, each of these studies employs different exposure assessment methodologies, which may account for some of the observed differences in the effect estimates (Faustini et al., 2015; Gan et al., 2017; F. Johnston et al., 2011; Morgan et al., 2010).

We also estimated a 2.0% (95% CI: 0.98 – 1.06) increase in the odds of same-day mortality among ages 65-84, and a 0% (95% CI: 0.96 – 1.05) increase in the odds of same-day mortality among ages 85 and over, on wildfire smoke days compared to non-wildfire smoke days, controlling for humidex. No studies report age group specific estimates for all non-traumatic mortality, with the exception of Kollanus et al. They estimate a 2.4% (95% CI: -6.5 – 12.0%) increase per 10μg/m^3 $PM_{2.5}$ in same-day non-accidental mortality among individuals 65 and over (Kollanus et al., 2016). While the effect estimate is similar, the confidence interval reported by Kollanus et al. suggests a smaller sample size, and the exposure metric used was substantially different. Thus, it is not clear whether these results are in line with our findings.

With respect to a lagged effect, we observe a 1.3% (95% CI: 1.00 – 1.02) increase in the odds of next day non-traumatic mortality for previous day wildfire smoke exposure versus non-wildfire smoke exposure, controlling for humidex and exposure for the four days prior to death. Other studies report evidence for a lagged effect of non-traumatic mortality from wildfire smoke exposure (Faustini et al., 2015; F. Johnston et al., 2011). Faustini et al. reported a 6.29% (95% CI: 1.00 – 11.85%) increase in cardiovascular mortality associated with cumulative lagged exposure from the five days prior to death on wildfire smoke days compared to non-wildfire smoke days (Faustini et al., 2015), while Johnston et al. (2011) observed a 5.0% (95% CI: 1.0 – 1.1) increase in the odds of non-accidental mortality due to prior day wildfire smoke exposure compared to non-wildfire smoke exposure. While these effect estimates are in the same direction as we report, the exposure assessment methodology used by Faustini et al. reports the change in mortality associated with cumulative lagged exposure for the five days prior to death, rather than reporting the change in mortality associated with exposure on the day prior to death, and thus cannot be directly compared.

In this analysis, we observed a 14.0% (95% CI: 1.01 – 1.29) increase in the odds of COPD mortality on wildfire smoke days compared to non-wildfire smoke days. Populations with underlying health conditions, and in particular, asthma and COPD, have been found to be more susceptible to wildfire smoke exposure compared to healthy populations in several studies examining hospital and ED admissions (Delfino et al., 2009; F. H. Johnston et al., 2014; Morgan et al., 2010; Colleen E. Reid et al., 2016), although no studies have examined the association between wildfire smoke exposure and COPD mortality. Delfino et al. found a 6.9% (95% CI: 0.9 - 13.1%) increase in COPD hospital admissions for ages 20-64 in an event-based study in southern California (Delfino et al., 2009). Johnston et al. reported a 12.0% same day increase in the odds of COPD-related ED admissions (95% CI: 1.02 – 1.24) in a time-series analysis of wildfire smoke events in Sydney (F. H. Johnston et al., 2014). Morgan et al. observed a 3.80% (95% CI: 1.40 – 6.26%) increase in COPD hospital admissions at lag day 2 in a time series analysis in Sydney (Morgan et al., 2010). Reid et al. reported a 2.0% increase in the risk of ED visits for COPD (95% CI 1.01 – 1.04) per 5 $\mu g/m^3$ increase in $PM_{2.5}$. The findings in this analysis are generally consistent with the results seen in the studies discussed above. However, these effects have predominantly been explored with respect to morbidity rather than mortality and cannot be directly compared to our results. Further research into the association between wildfire smoke exposure and COPD mortality should be conducted to confirm this observation.

Among individuals ages 45-64, a 35.0% increase in the odds of same-day respiratory mortality was found (95% CI: 1.09 – 1.67). This finding is not seen in the literature, where respiratory hospital and ED admissions associated with wildfire smoke are more pronounced in older adults (ages ≥ 65) compared to adults under 65 (Delfino et al., 2009; Morgan et al., 2010; Wettstein et al., 2018). Only one study examines respiratory mortality by age due to small subgroup sample size in several studies (Linares et al., 2015). They report no association with respiratory mortality among all ages, but report a 6.1% (95% CI: 0.99 – 1.14) increase in risk of respiratory mortality among adults 75 and over for wildfire smoke exposure two days prior to death (Linares et al., 2015). We observed an 8.0% (95% CI: 0.96 – 1.21) increase in the odds of respiratory mortality and a 14.0% (95% CI: 0.99 – 1.31) increase in the odds of COPD mortality among adults ages 65-84. Our evidence of an increased risk among individuals ages 45-64 may suggest a difference in effect on working populations versus non-working populations in our region. Older working populations, ages 45-64, may be more likely to be exposed to wildfire smoke than older non-working populations ages 65 and over, as they may be more likely to go outside in order to travel to work. Further, this population is less healthy than younger working populations under 45, who are rarely diagnosed with COPD (Arshukova, Dobretsova, & Demko, 2015; Fletcher et al., 2011). Additional research should be conducted to further examine the risk to this population.

This analysis is limited by the challenges of separating anthropogenic $PM_{2.5}$ and wildfire smoke $PM_{2.5}$. Significant effort was put into determining a viable threshold, but it is likely some misclassification exists. The sensitivity analysis conducted using 20.4 $\mu g/m^3$ as the wildfire smoke threshold (Table B1) shows the effects from wildfire smoke are sensitive to the wildfire smoke threshold, and that the effect seen in this study may be capturing both the risk of anthropogenic $PM_{2.5}$ and wildfire smoke $PM_{2.5}$. Further, existing wildfire smoke and mortality studies employ a wide variety of exposure assessment methodologies, impeding accurate comparison of effect estimates across studies. Additional research is needed to develop more accurate methodology in the estimation of wildfire smoke affected days.

Additional limitations include the assumption that $PM_{2.5}$ concentrations at each monitor represent the exposure for the entire population attributed to each monitor. While these concentrations have been checked against the modeled $PM_{2.5}$ at each 4x4km grid, the concentrations still rely on monitor performance. Additionally, the air quality monitors do not necessarily represent personal exposure for each grid cell, resulting in potential exposure misclassification.

Washington State geocoded mortality data may not necessarily correspond to the location of exposure. The geocoded location of death was used to link each individual death to a grid cell is determined by the location of death. If the location of death was not the location of exposure, this may result in misclassification of exposure, although this is less likely for same day exposures. Further, misclassification of cause of death on death certificates may lead to improper stratification of the cause of death results.

For estimates of non-traumatic mortality stratified by median household income, an ecological indicator was employed, assigning individuals the census tract median household income. This metric misclassifies individual income, but serves as a proxy for neighborhood level socioeconomic status. Due to inadequate median household income data for 2006-2009, estimates for that time period are prone to higher rates of misclassification. Further analyses should develop more accurate proxies for individual or neighborhood level socioeconomic status.

Conclusion

This study is the first to estimate mortality risk associated with wildfire smoke exposure in Washington State. Similar work has been explored in California and British Columbia. This study uses a tiered approach to exposure assessment, minimizing the false positive anthropogenic-related days in urban areas and false negative days in rural areas. This work will inform local and state risk communication efforts and decision-making during future wildfire smoke events, especially for the susceptible subpopulations identified in this study: those with COPD and older working populations with underlying respiratory conditions. Additional research is needed in Washington State to better characterize the association between wildfire smoke exposure and less severe health endpoints of interest, including hospitalizations and ED visits.

Aim 3: Communication of findings to partner agencies, conferences, and to the public

Findings from this research study will be presented at the Washington State Environmental Health Association (WSEHA) conference in May 2019 in Yakima, Washington. The audience of this conference includes environmental health professionals, academics, and students from across the state. The goal of this presentation will be to inform audience members of my work and to increase awareness to inform risk communication efforts for wildfire smoke events, and to aid in prioritizing populations for wildfire smoke relief programs.

Findings from this research study will also be presented at the National Environmental Health Association (NEHA) conference in July 2019 in Nashville, Tennessee. The audience of this conference will include environmental health professionals, academics, and students from across the U.S. The goal of this presentation will be to inform audience members of my work and to increase knowledge of risk association with wildfire smoke exposure, and how this work may be applied to other settings and can be used to inform risk communication.

Findings from this research study will presented in a variety of state and local venues. The findings will be used to inform risk communication efforts at the state Department of Health (DOH) and Department of Ecology (ECY). Additional presentations will be made to staff at Puget Sound Clean Air Agency (PSCAA), staff in the EPA Region 10 office, and to a group of emergency managers in King County (PHSKC).

Conclusion

The main goal of this research study was to determine a suitable wildfire smoke definition and to examine the association between wildfire smoke exposure in Washington State during wildfire season 2006-2017 and non-traumatic mortality. The study classified wildfire smoke definitions according to external metrics. The study also found evidence for an association between wildfire smoke exposure and increased odds of COPD mortality and respiratory mortality among ages 45-64 in particular, but did not find evidence for an increase in the odds of all-non traumatic mortality. Further research is needed to improve exposure methods for ease of comparison across studies. Additionally, further research should examine specific causes of mortality due to wildfire smoke exposure in other areas in the U.S., and should examine the potential association between wildfire smoke exposure in Washington and increased risk of hospitalizations to better understand the public health burden.

This study provides baseline effect estimates to inform risk communication efforts for state and local health officials in Washington State for future wildfire smoke events. This study also provides evidence for targeting messaging, interventions, and resources towards particular populations, including all individuals with COPD and pre-existing respiratory and cardiovascular, conditions, and not just those 65 and over. Further research should be conducted to better quantify the morbidity burden associated with wildfire smoke exposure in Washington State.

References

Abatzoglou, J. T., & Williams, A. P. (2016). Impact of anthropogenic climate change on wildfire across western US forests. https://doi.org/10.1073/pnas.1607171113

Alman, B. L., Pfister, G., Hao, H., Stowell, J., Hu, X., Liu, Y., & Strickland, M. J. (2016). The association of wildfire smoke with respiratory and cardiovascular emergency department visits in Colorado in 2012: A case crossover study. *Environmental Health: A Global Access Science Source*, *15*(1), 1–9. https://doi.org/10.1186/s12940-016-0146-8

Alves, C. A., Vicente, A., Monteiro, C., Gonçalves, C., Evtyugina, M., & Pio, C. (2011). Science of the Total Environment Emission of trace gases and organic components in smoke particles from a wild fi re in a mixed-evergreen forest in Portugal. *Science of the Total Environment, The*, *409*(8), 1466–1475. https://doi.org/10.1016/j.scitotenv.2010.12.025

Ames, R. B., Fox, D. G., Malm, W. C., Schichtel, B. A., & Service, N. P. (n.d.). Preliminary apportionments of carbonaceous aerosols to wild fire smoke using observations from the IMPROVE network, (76), 1–10.

Analitis, A., Georgiadis, I., & Katsouyanni, K. (2012). Forest fires are associated with elevated mortality in a dense urban setting. *Occupational and Environmental Medicine*, *69*(3), 158–162. https://doi.org/10.1136/oem.2010.064238

Arshukova, I. L., Dobretsova, E. A., & Demko, I. V. (2015). Epidemiology of chronic obstructive pulmonary disease : a population-based study in Krasnoyarsk region , Russia, 1781–1786.

Black, C., Tesfaigzi, Y., Bassein, J. A., & Miller, L. A. (2017). Wildfire smoke exposure and human health: Significant gaps in research for a growing public health issue. *Environmental Toxicology and Pharmacology*, *55*(August), 186–195. https://doi.org/10.1016/j.etap.2017.08.022

Cascio, W. E. (2018). Wildland fire smoke and human health. *Science of the Total Environment*, *624*, 586–595.

Delfino, R. J., Brummel, S., Wu, J., Stern, H., Ostro, B., Lipsett, M., … Gillen, D. L. (2009). The relationship of respiratory and cardiovascular hospital admissions to the southern California wildfires of 2003 The relationship of respiratory and cardiovascular hospital admissions to the southern California wildfires of 2003, *66*(Mar), 1–24. https://doi.org/10.1136/oem.2008.041376

Dell, K. O., Ford, B., Fischer, E. V, & Pierce, R. (2019). Contribution of Wildland-Fire Smoke to US PM 2.5 and Its In fl uence on Recent Trends. https://doi.org/10.1021/acs.est.8b05430

Dennekamp, M., Straney, L. D., Erbas, B., Abramson, M. J., Keywood, M., Smith, K., … Tonkin, A. M. (2015). Forest Fire Smoke Exposures and Out-of-Hospital Cardiac Arrests, *123*(10), 959–964. https://doi.org/10.1289/ehp.1408436

Dennison, P. E., Brewer, S. C., Arnold, J. D., & Moritz, M. a. (2014). Geophysical Research Letters. *Geophysical Prospecting*, *41*, 2928–2933. https://doi.org/10.1002/2014GL061184.Received

Dockery, D., Pope, C., Xu, X., Spengler, J., Ware, J., Fay, M., … Speizer, F. (1993). An Association between Air Pollution and Mortality in Six U.S. Cities. *N Engl J Med*, *329*, 1753–1759.

Drosatou, A. D., Skyllakou, K., Theodoritsi, G. N., & Pandis, S. N. (2019). Positive matrix factorization of organic aerosol : insights from a chemical transport model, 973–986.

Elliott, C. T., Henderson, S. B., & Wan, V. (2013). Time series analysis of fine particulate matter and asthma reliever dispensations in populations affected by forest fires. *Environmental Health: A Global Access Science Source*, *12*(1), 1–9. https://doi.org/10.1186/1476-069X-12-11

Faustini, A., Alessandrini, E. R., Pey, J., Perez, N., Samoli, E., Querol, X., … Pascal, M. (2015). Short-term effects of particulate matter on mortality during forest fires in Southern Europe: Results of the MED-PARTICLES project. *Occupational and Environmental Medicine*, *72*(5), 323–329. https://doi.org/10.1136/oemed-2014-102459

Fletcher, M. J., Upton, J., Taylor-fishwick, J., Buist, S. A., Jenkins, C., Hutton, J., … Walker, S. (2011). COPD uncovered : an international survey on the impact of chronic obstructive pulmonary disease [COPD] on a working age population, 1–13.

Gan, R. W., Ford, B., Lassman, W., Pfister, G., Vaidyanathan, A., Fischer, E., … Magzamen, S. (2017). associations with cardiopulmonary-related hospital admissions, *1*(3), 122–136. https://doi.org/10.1002/2017GH000073.Comparison

Grell, G. A., Peckham, S. E., Schmitz, R., Mckeen, S. A., Frost, G., Skamarock, W. C., & Eder, B. (2005). Fully coupled '" online "' chemistry within the WRF model, *39*, 6957–6975. https://doi.org/10.1016/j.atmosenv.2005.04.027

Haikerwal, A., Akram, M., Monaco, A. Del, Smith, K., Sim, M. R., Meyer, M., … Dennekamp, M. (2015). Impact of fine particulate matter (PM2.5) exposure during wildfires on cardiovascular health outcomes. *Journal of the American Heart Association*, *4*(7), 1–10. https://doi.org/10.1161/JAHA.114.001653

Hand, J. L., Schichtel, B. A., Malm, W. C., & Frank, N. H. (2013). Spatial and Temporal Trends in PM 2.5 Organic and Elemental Carbon across the United States. *Advances in Meteorology*, *2013*.

Hänninen, O. O., Salonen, R. O., Koistinen, K., Lanki, T., Barregard, L., & Jantunen, M. (2009). Population exposure to fine particles and estimated excess mortality in Finland from an East European wildfire episode. *Journal of Exposure Science and Environmental Epidemiology*, *19*(4), 414–422. https://doi.org/10.1038/jes.2008.31

Henderson, S. B., Brauer, M., Macnab, Y. C., Kennedy, S. M., Environmental, S., Perspectives, H., … Kennedy, S. M. (2011). Three Measures of Forest Fire Smoke Exposure and Their Associations with Respiratory and Cardiovascular Health Outcomes in a Population-Based Cohort Published by : The National Institute of Environmental Health Sciences Stable URL : http://www.jstor.org/s. *Environmental Health Perspectives*, *119*(9), 1266–1271.

Hohner, A. K., Cawley, K., Oropeza, J., Summers, R. S., & Rosario-ortiz, F. L. (2016). Drinking water treatment response following a Colorado wild fi re. *Water Research*, *105*, 187–198.

https://doi.org/10.1016/j.watres.2016.08.034

Hong, K. Y., King, G. H., Saraswat, A., & Henderson, S. B. (2017). Seasonal ambient particulate matter and population health outcomes among communities impacted by road dust in British Columbia, Canada. *Journal of the Air and Waste Management Association*, *67*(9), 986–999. https://doi.org/10.1080/10962247.2017.1315348

Interagency Working Group on Climate Change and Health. (2010). *A Human Health Perspective on Climate Change: A Report Outlining the Research Needs on the Human Health Effects of Climate Change*.

Jaffe, D., Hafner, W., Chand, D., Westerling, A., & Spracklen, D. (2008). Interannual Variations in PM2.5 due to Wildfires in the Western United. *Environ Sci Technol*, *42*(8), 2812–2818.

Janes, H., Sheppard, L., & Lumley, T. (2005). Case-crossover analyses of air pollution exposure data: Referent selection strategies and their implications for bias. *Epidemiology*, *16*(6), 717–726. https://doi.org/10.1097/01.ede.0000181315.18836.9d

Johnston, F. H., Purdie, S., Jalaludin, B., Martin, K. L., Henderson, S. B., & Morgan, G. G. (2014). Air pollution events from forest fires and emergency department attendances in Sydney, Australia 1996-2007: A case-crossover analysis. *Environmental Health: A Global Access Science Source*, *13*(1), 1–9. https://doi.org/10.1186/1476-069X-13-105

Johnston, F., Hanigan, I., Henderson, S., Morgan, G., & Bowman, D. (2011). Extreme air pollution events from bushfires and dust storms and their association with mortality in Sydney, Australia 1994-2007. *Environmental Research*, *111*(6), 811–816. https://doi.org/10.1016/j.envres.2011.05.007

Joyce, L., Running, S., Breshears, D., Dale, V., Malmsheimer, R., Sampson, R., … Woodall, C. (2014). *Ch. 7: Forests. Climate Change Iimpacts in the United States: The Third National Climate Assessment*. https://doi.org/10.7930/J0Z60KZC.On

Kang, C.-M., Gold, D., & Koutrakis, P. (2014). Downwind O3 and PM2.5 speciation during the wildfires in 2002 and 2010. *Atmospheric Environment*, *95*, 511–519.

Kollanus, V., Prank, M., Gens, A., Soares, J., Vira, J., Kukkonen, J., & Sofiev, M. (2017). Mortality due to Vegetation Fire – Originated PM 2 . 5 Exposure in, (1), 30–37.

Kollanus, V., Tiittanen, P., Niemi, J. V., & Lanki, T. (2016). Effects of long-range transported air pollution from vegetation fires on daily mortality and hospital admissions in the Helsinki metropolitan area, Finland. *Environmental Research*, *151*, 351–358. https://doi.org/10.1016/j.envres.2016.08.003

Lassman, W., Ford, B., Gan, R. W., Fischer, E. V, & Pierce, J. R. (2017). Spatial and temporal estimates of population exposure to wild fi re smoke during the Washington state 2012 wild fi re season using blended model, satellite, and in situ data. https://doi.org/10.1002/2017GH000049

Li, M. H., Fan, L. C., Mao, B., Yang, J. W., Choi, A. M. K., Cao, W. J., & Xu, J. F. (2016). Short-term exposure to ambient fine particulate matter increases hospitalizations and mortality in COPD: A systematic review and meta-analysis. *Chest*, *149*(2), 447–458. https://doi.org/10.1378/chest.15-0513

Linares, C., Carmona, R., Salvador, P., & Díaz, J. (2018). Impact on mortality of biomass combustion from wildfires in Spain: A regional analysis. *Science of the Total Environment*, *622–623*, 547–555. https://doi.org/10.1016/j.scitotenv.2017.11.321

Linares, C., Carmona, R., Tobías, A., Mirón, I. J., & Díaz, J. (2015). Influence of advections of particulate matter from biomass combustion on specific-cause mortality in Madrid in the period 2004–2009. *Environmental Science and Pollution Research*, *22*(9), 7012–7019. https://doi.org/10.1007/s11356-014-3916-2

Littell, J., McKenzie, D., Peterson, D., & Westerling, A. (2009). Climate and wildfire area burned in western U.S. *Ecological Applications*, *19*(4), 1003–1021.

Liu, J. C., & Peng, R. (2018). The impact of wild fi re smoke on compositions of fi ne particulate matter by ecoregion in the Western US. https://doi.org/10.1038/s41370-018-0064-7

Liu, J. C., Pereira, G., Uhl, S. A., Bravo, M. A., & Bell, M. L. (2015). A systematic review of the physical health impacts from non-occupational exposure to wild fi re smoke, *136*, 120–132. https://doi.org/10.1016/j.envres.2014.10.015

Liu, J. C., Wilson, A., Mickley, L. J., Dominici, F., Ebisu, K., Wang, Y., … Bell, M. L. (2017). Wildfire-specific fine particulate matter and risk of hospital admissions in urban and rural counties. *Epidemiology*, *28*(1), 77–85. https://doi.org/10.1097/EDE.0000000000000556

Lumley, T., & Levy, D. (2000). Bias in the case-crossover design: implications for studies of air pollution. *Environmetrics*, *11*, 689–704. https://doi.org/10.1002/1099-095X(200011/12)11:6<689::AID-ENV439>3.0.CO;2-N

Mallia, D., Lin, J., Urbanksi, S., Ehleringer, J., & Nehrkorn, T. (2014). Impacts of upwind wildfire emissions on CO, CO2, and PM2.5 concentrations in Salt Lake City, Utah. *Journal of Geophysical Research: Atmospheres*, *120*, 147–166. https://doi.org/10.1002/2014JD022472.Received

Malm, W. C., Schichtel, B. A., & Pitchford, M. L. (2011). Uncertainties in PM 2 . 5 Gravimetric and Speciation Measurements and What We Can Learn from Them, *61*(November), 1131–1149. https://doi.org/10.1080/10473289.2011.603998

Management, O. of F. (2017). Population & demographics. Retrieved from https://www.ofm.wa.gov/washington-data-research/population-demographics

Martin, K. L., Hanigan, I. C., Morgan, G. G., Henderson, S. B., & Johnston, F. H. (2013). Air pollution from bushfires and their association with hospital admissions in Sydney, Newcastle and Wollongong, Australia 1994-2007. *Australian and New Zealand Journal of Public Health*, *37*(3), 238–243. https://doi.org/10.1111/1753-6405.12065

Masterton, J., & Richardson, F. (1979). *Humidex: a method of quantifying human discomfort due to excessive heat and humidity*. Downsview, Ontario: Environment Canada, Atmospheric Environment.

Moeltner, K., Kim, M., Zhu, E., & Yang, W. (2013). Wildfire smoke and health impacts : A closer look at fire attributes and their marginal effects. *Journal of Environmental Economics and Management*, *66*(3), 476–496. https://doi.org/10.1016/j.jeem.2013.09.004

Morgan, G., Sheppeard, V., Khalaj, B., Ayyar, A., Lincoln, D., Jalaludin, B., … Lumley, T. (2010). Effects of bushfire smoke on daily mortality and hospital admissions in Sydney, Australia. *Epidemiology*, *21*(1), 47–55.

Mott, J. A., Mannino, D. M., Alverson, C. J., Kiyu, A., Hashim, J., Lee, T., … Redd, S. C. (2005). Cardiorespiratory hospitalizations associated with smoke exposure during the 1997 Southeast Asian forest fires. *International Journal of Hygiene and Environmental Health*, *208*(1–2), 75–85. https://doi.org/10.1016/j.ijheh.2005.01.018

Munoz-Alpizar, R., Pavlovic, R., Moran, M. D., Chen, J., Gravel, S., Henderson, S. B., … Bouchet, V. (2016). Exposure over North America as Determined from Operational Air Quality Forecasts. https://doi.org/10.3390/atmos8090179

Naeher, L. P., Brauer, M., Lipsett, M., Zelikoff, J. T., Simpson, C. D., Koenig, J. Q., & Smith, K. R. (2010). Woodsmoke Health Effects : A Review, *8378*. https://doi.org/10.1080/08958370600985875

National Oceanic and Atmospheric Administration. (2018). Hazard Mapping System Fire and Smoke Product. Retrieved from https://www.ospo.noaa.gov/Products/land/hms.html

Orellano, P., Quaranta, N., Reynoso, J., Balbi, B., & Vasquez, J. (2017). Effect of outdoor air pollution on asthma exacerbations in children and adults: Systematic review and multilevel meta-analysis. *PLoS ONE*, *12*(3), 1–15. https://doi.org/10.1371/journal.pone.0174050

Parry, M. L. (2007). *Technical Summary. Climate Change 2007: Impacts, Adaptation, and Vulnerability*. Cambridge University Press, Cambridge, UK.

Pilz, J., & Spöck, G. (2008). Why do we need and how should we implement Bayesian kriging methods. *Stochastic Environmental Research and Risk Assessment*, *22*(5), 621–632. https://doi.org/10.1007/s00477-007-0165-7

R Development Core Team. (2018). R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from http://www.r-project.org

Ragan, M. (2018). *2017 Ambient Air Monitoring Network Report*. Retrieved from https://fortress.wa.gov/ecy/publications/SummaryPages/1602001.html

Reid, C. E., Brauer, M., Johnston, F. H., Jerrett, M., Balmes, J. R., & Elliott, C. T. (2016). Critical review of health impacts of wildfire smoke exposure. *Environmental Health Perspectives*, *124*(9), 1334–1343. https://doi.org/10.1289/ehp.1409277

Reid, C. E., Jerrett, M., Tager, I. B., Petersen, M. L., Mann, J. K., & Balmes, J. R. (2016). Differential respiratory health effects from the 2008 northern California wildfires: A spatiotemporal approach. *Environmental Research*, *150*, 227–235. https://doi.org/10.1016/j.envres.2016.06.012

Sastry, N. (2002). Forest fires, air pollution, and mortality in southeast Asia. *Demography Feb*, *39*(1), 1–23.

Shaposhnikov, D., Revich, B., Bellander, T., Bedada, G. B., Bottai, M., Kharkova, T., … Pershagen, G. (2014). Mortality related to air pollution with the Moscow heat wave and wildfire of 2010. *Epidemiology*, *25*(3), 359–364. https://doi.org/10.1097/EDE.0000000000000090

Spracklen, D. V, Logan, J. A., Mickley, L. J., Park, R. J., Yevich, R., Westerling, A. L., & Jaffe, D. A. (2007). Wildfires drive interannual variability of organic carbon aerosol in the western U . S . in summer, *34*, 2–5. https://doi.org/10.1029/2007GL030037

Tinling, M. A., West, J. J., Cascio, W. E., Kilaru, V., & Rappold, A. G. (2016). Repeating cardiopulmonary health effects in rural North Carolina population during a second large peat wildfire. *Environmental Health: A Global Access Science Source*, *15*(1), 1–12. https://doi.org/10.1186/s12940-016-0093-4

Turpin, B. J., Lim, H., Turpin, B. J., & Lim, H. (2010). Species Contributions to PM2 . 5 Mass Concentrations : Revisiting Common Assumptions for Estimating Organic Mass Species Contributions to PM2 . 5 Mass Concentrations : Revisiting Common Assumptions for Estimating Organic Mass, *6826*.

U. S. Environmental Protection Agency. (2010). Quantitative Health Risk Assessment for Particulate Matter. *U. S. Environmental Protection Agency*, 1–596. https://doi.org/10.1097/SLA.0b013e31818c7082

Urbanski, S., Hao, W., & Baker, S. (2008). Chapter 4: Chemical Composition of Wildland Fire Emissions. *Developments in Environmental Science*, *8*, 79–107. https://doi.org/https://doi.org/10.1016/S1474-8177(08)00004-1

US EPA. (n.d.). NAAQS Table. Retrieved from https://www.epa.gov/criteria-air-pollutants/naaqs-table

US EPA. (2009). *Integrated Science Assessment (ISA) for Particulate Matter*. Research Triangle Park, NC.

Vicente, A., Alves, C. A., Calvo, A. I., Fernandes, A. P., Nunes, T., Monteiro, C., … Pio, C. (2013). Emission factors and detailed chemical composition of smoke particles from the 2010 wildfire season. *Atmospheric Environment*, *71*, 295–303.

Vila, K., Nunes, R., Ignotti, E., & Hacon, S. D. S. (2013). Mortalidade por doenças circulatórias na população Idosa e Exposição un PM 2 , 5 em decorrência das Queimadas na Amazônia Brasileira em 2005 Mortalidad por diseases Cardiovasculares en los ancianos y la Exposición un PM 2 , 5 A consecuencia de la quema en, *29*(3), 589–598.

WA DOH; WA ECY. (n.d.). WAQA: Washington Air Quality Advisory. Retrieved from https://fortress.wa.gov/ecy/publications/documents/1802024.pdf

Washington State Department of Health. (2012). *Surveillance Investigation of the Cardiopulmonary Health Effects of the 2012 Wildfires in North Central Washington State*.

Washington State University. (2019). AIRPACT. Retrieved from http://lar.wsu.edu/airpact/index.html

Westerling, A. L. (2016). (2016). Increasing western US forest wildfire activity: sensitivity to changes in the timing of spring. Philos. Trans. Royal Soc. B, 371, Pii: 20160373.

Wettstein, Z. S., Hoshiko, S., Fahimi, J., Harrison, R. J., Cascio, W. E., & Rappold, A. G. (2018). Cardiovascular and Cerebrovascular Emergency Department Visits Associated With Wildfire Smoke Exposure in California in 2015. *Journal of the American Heart Association*, *7*(8), e007492. https://doi.org/10.1161/JAHA.117.007492

Xie, M., Hannigan, M. P., Dutton, S. J., Milford, J. B., Hemann, J. G., Miller, S. L., … Vedal, S. (2012). Positive Matrix Factorization of PM 2.5 : Comparison and Implications of Using Di ff erent Speciation Data Sets. https://doi.org/10.1021/es302358g

Yao, J., Eyamie, J., & Henderson, S. B. (2016). Evaluation of a spatially resolved forest fire smoke model for population-based epidemiologic exposure assessment. *Journal of Exposure Science and Environmental Epidemiology*, *26*(3), 233–240. https://doi.org/10.1038/jes.2014.67

Yao, J., & Henderson, S. B. (2014). An empirical model to estimate daily forest fire smoke exposure over a large geographic area using air quality, meteorological, and remote sensing data. *Journal of Exposure Science and Environmental Epidemiology*, *24*(3), 328–335. https://doi.org/10.1038/jes.2013.87

Youssouf, H., Liousse, C., Roblou, L., Assamoi, E. M., Salonen, R. O., Maesano, C., … Annesi-Maesano, I. (2014). Non-accidental health impacts of wildfire smoke. *International Journal of Environmental Research and Public Health*, *11*(11), 11772–11804. https://doi.org/10.3390/ijerph111111772

Zu, K., Tao, G., Long, C., Goodman, J., & Valberg, P. (2016). Long-range fine particulate matter from the 2002 Quebec forest fires and daily mortality in Greater Boston and New York City. *Air Quality, Atmosphere and Health*, *9*(3), 213–221. https://doi.org/10.1007/s11869-015-0332-9

Appendix A. Aim 1 additional plots

Figure A1. Organic Carbon distribution by smoke day method

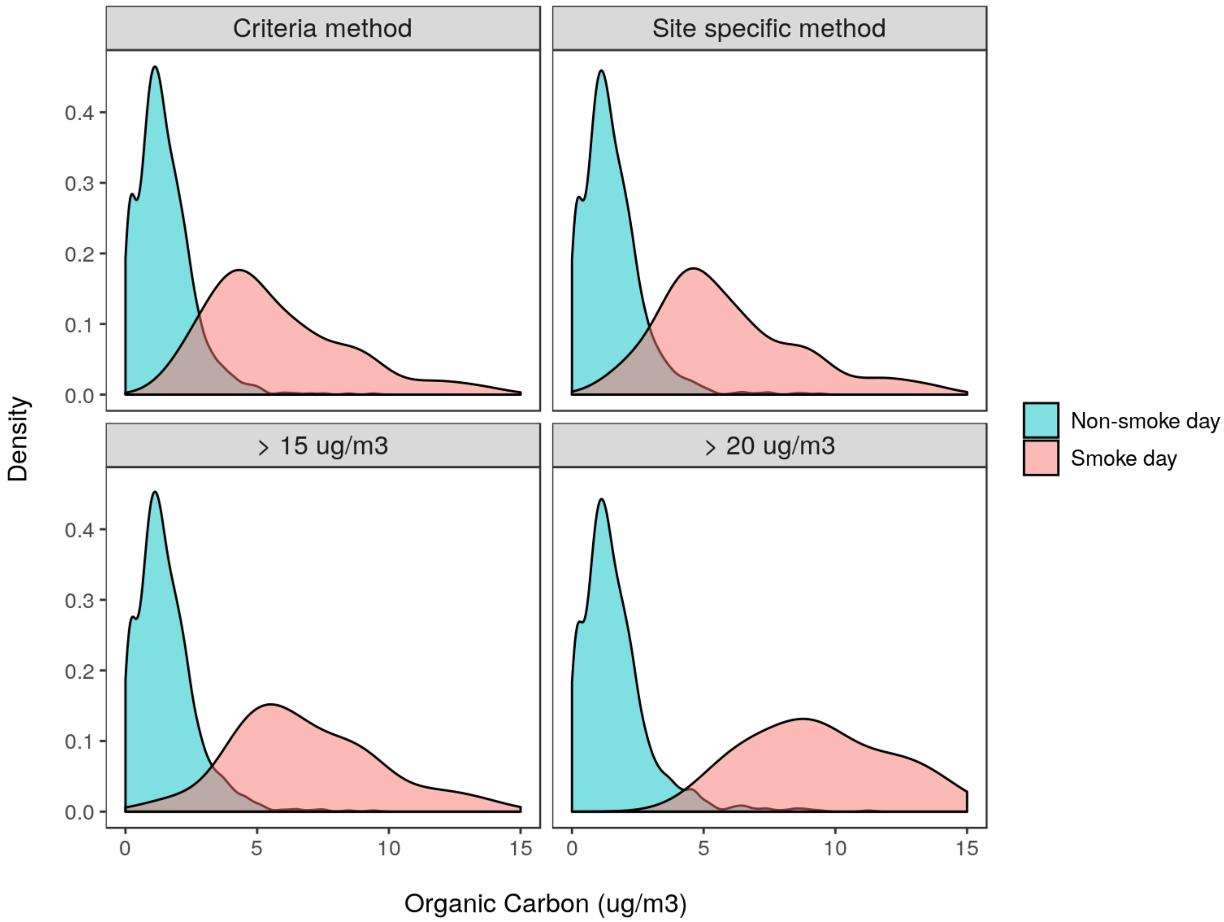


Figure A2. $PM_{2.5}$ distribution by smoke day method

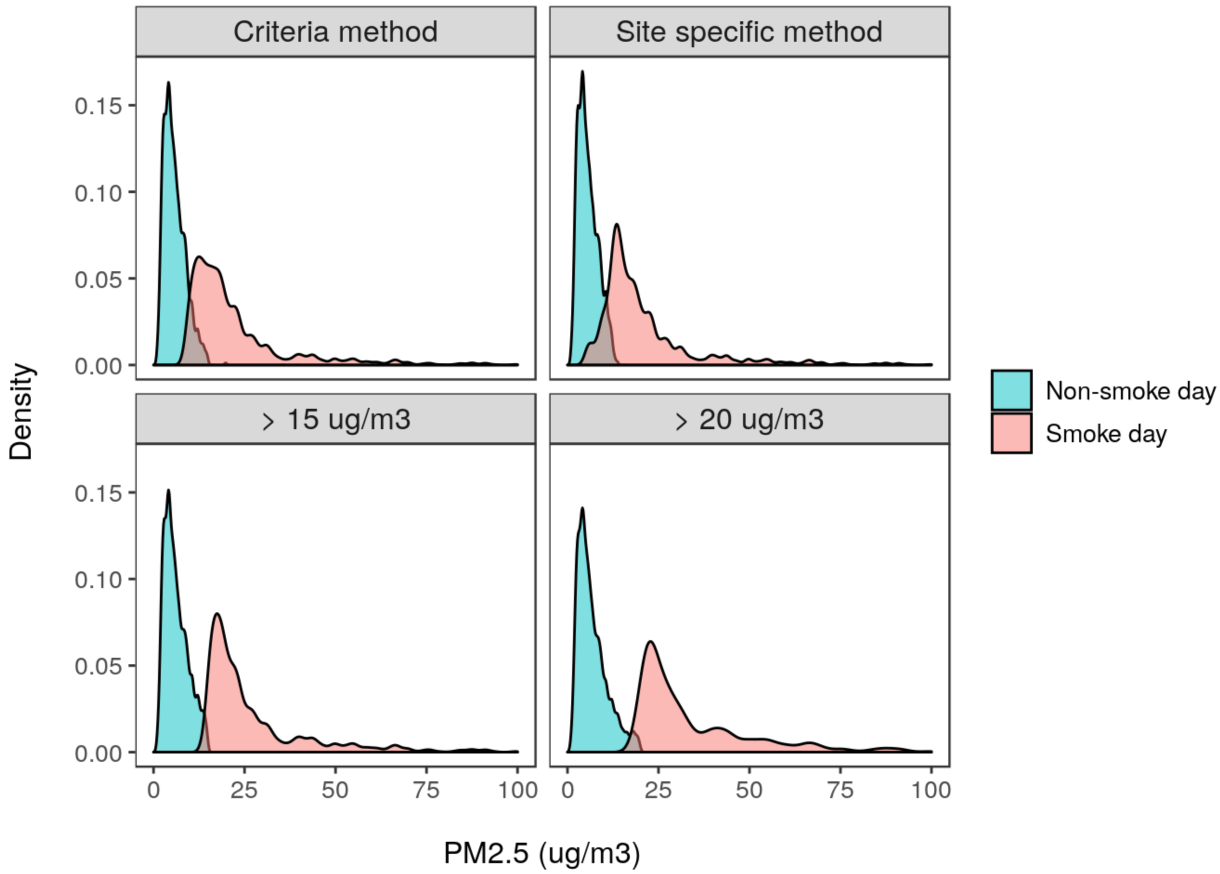


Figure A3. Count of days with CSN or IMPROVE data, by site and year

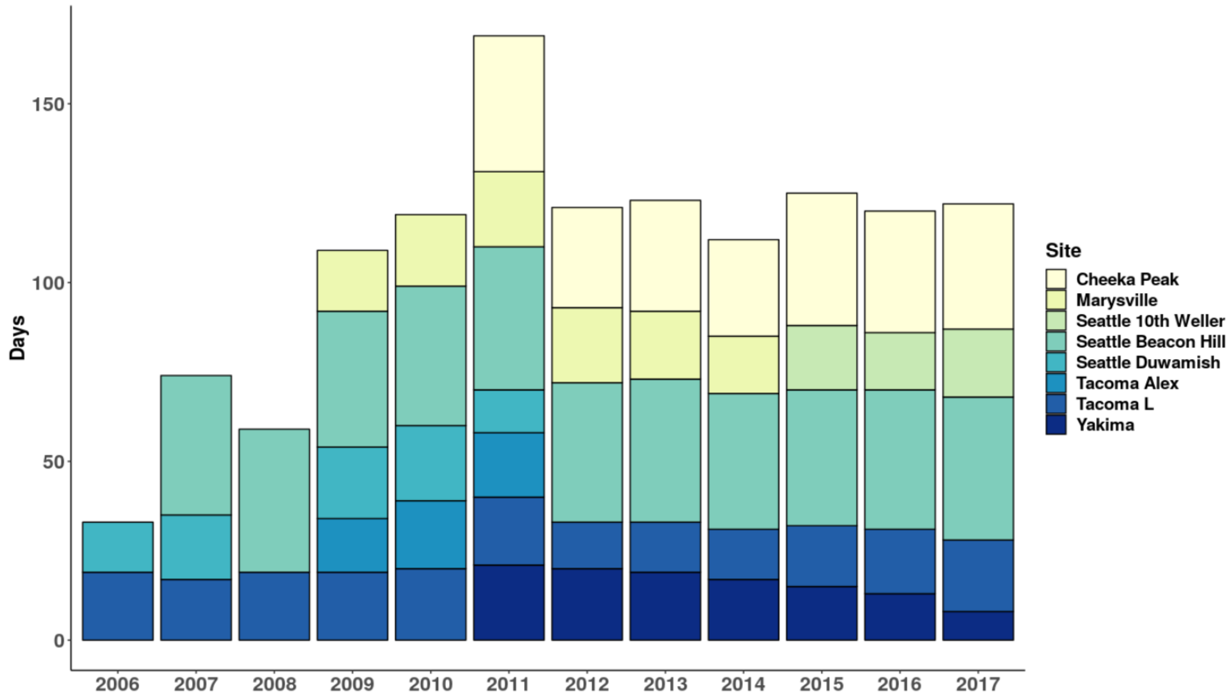


Figure A4. $PM_{2.5}$ distribution by Gan et al. smoke method

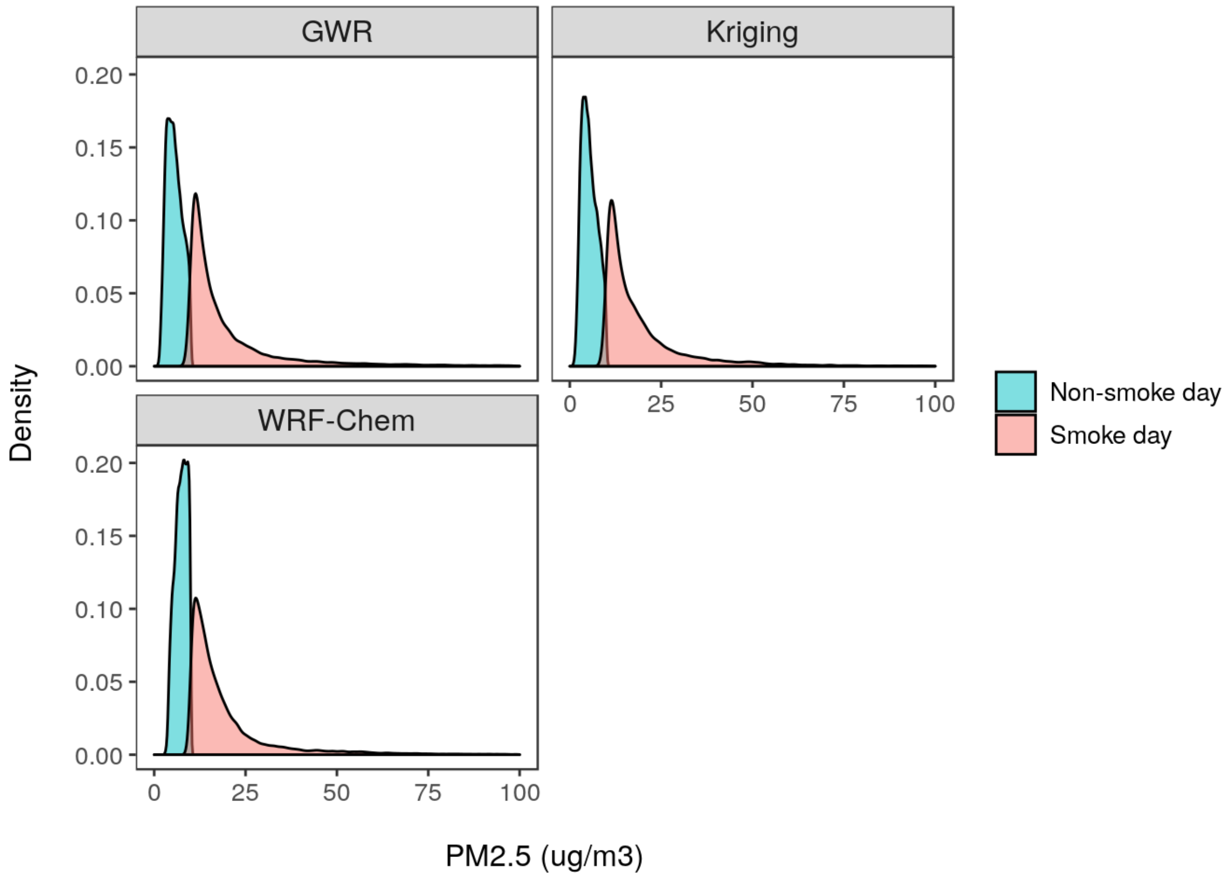


Figure A5. Percent of wildfire smoke grid cell days, by method, classified as wildfire smoke days by Gan et al., by method

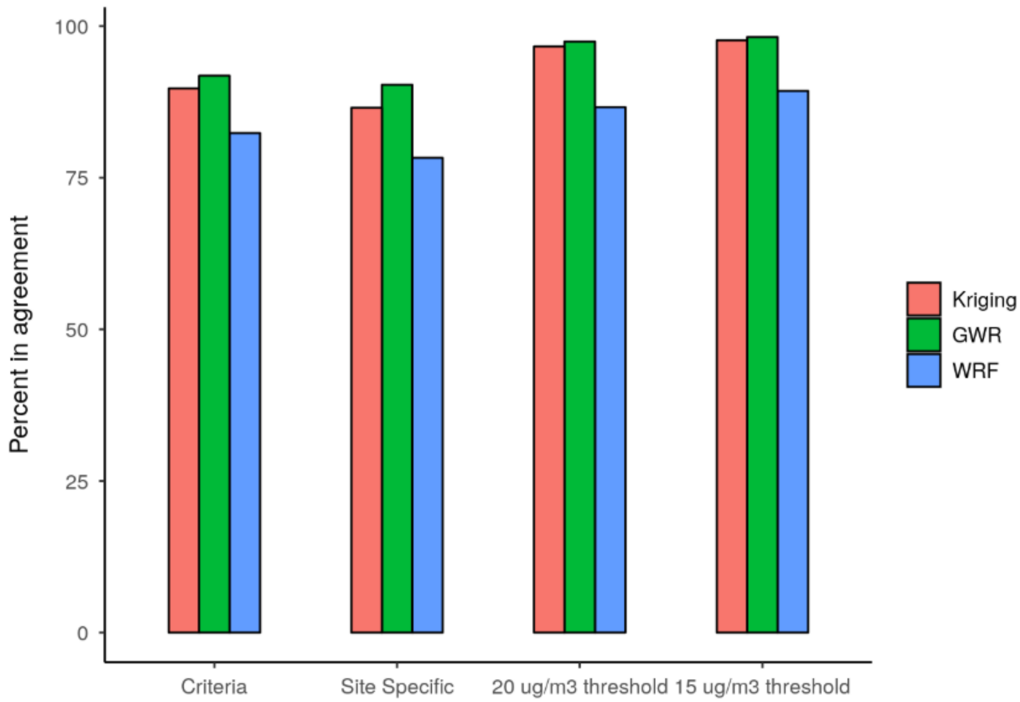


Figure A6. Percent of non-wildfire smoke grid cell days, by method, classified as non-wildfire smoke days by Gan et al., by method

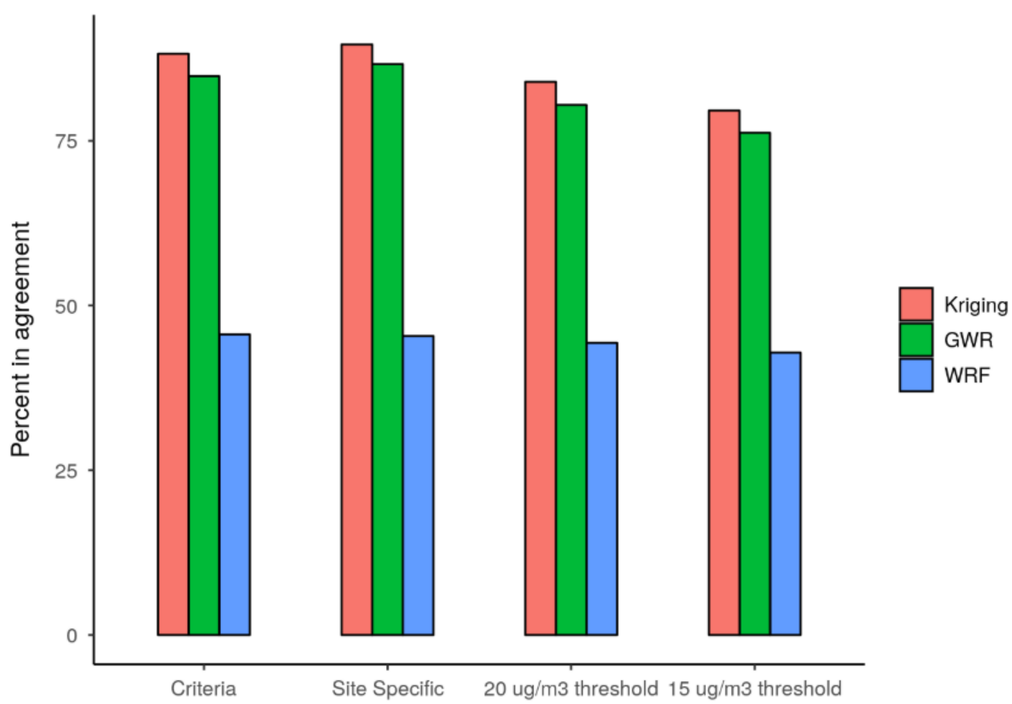


Figure A7. $PM_{2.5}$ distribution of criteria method wildfire smoke and non-wildfire smoke days, misclassified by Gan et al. method

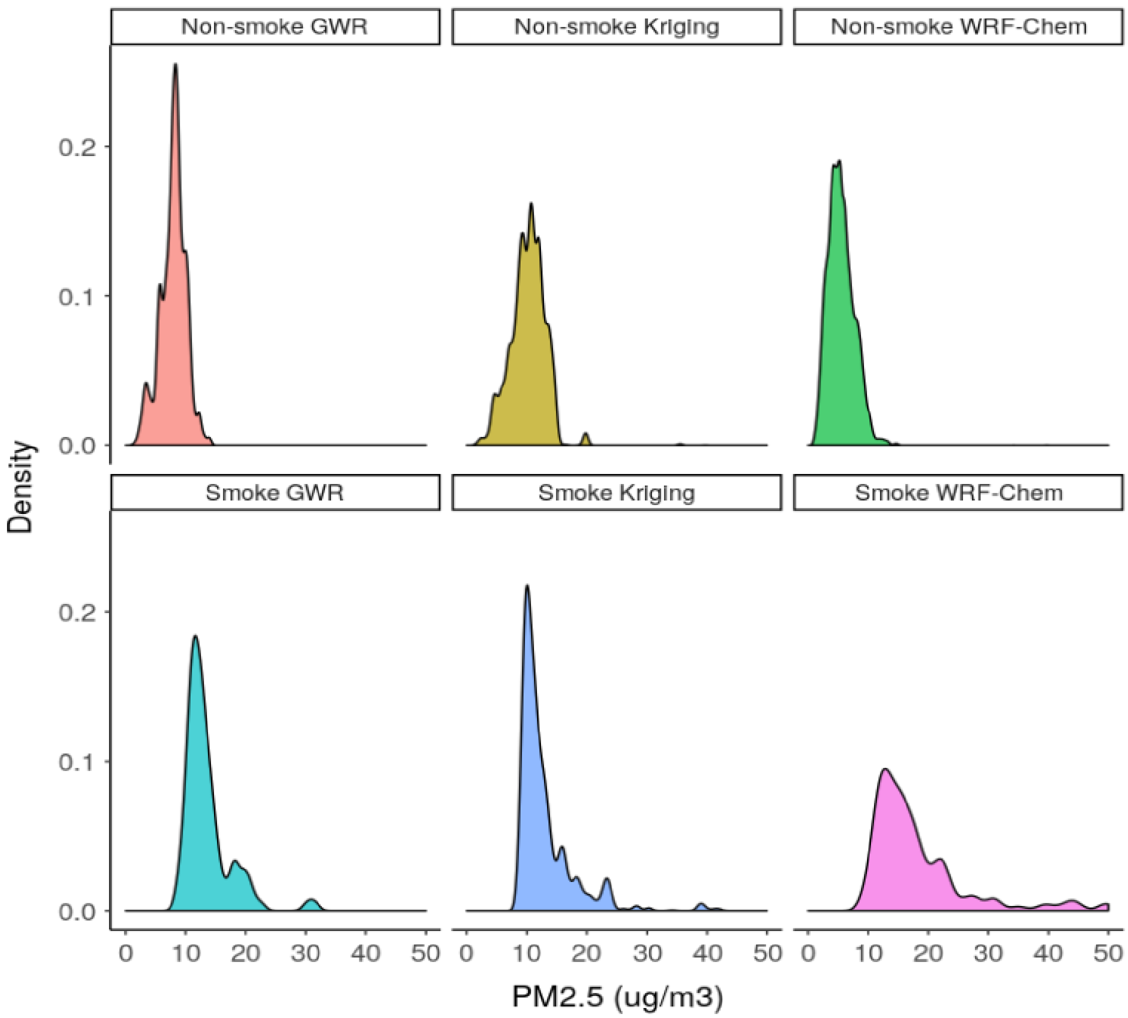


Figure A8. $PM_{2.5}$ distribution of site-specific method wildfire smoke and non-wildfire smoke days, misclassified by Gan et al. method

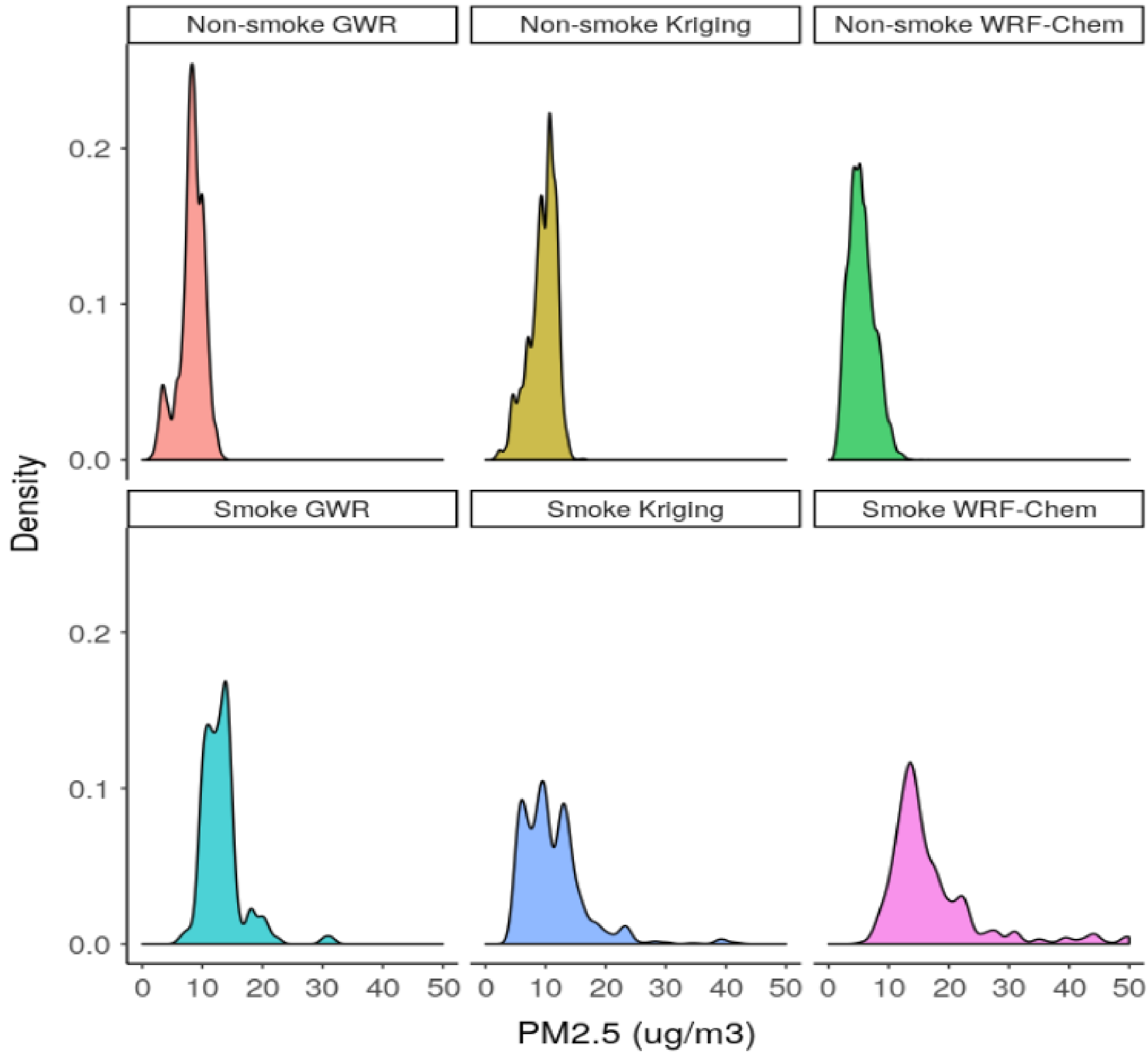


Figure A9. $PM_{2.5}$ distribution of 15 μg/m^3 threshold method wildfire smoke and non-wildfire smoke days, misclassified by Gan et al. method

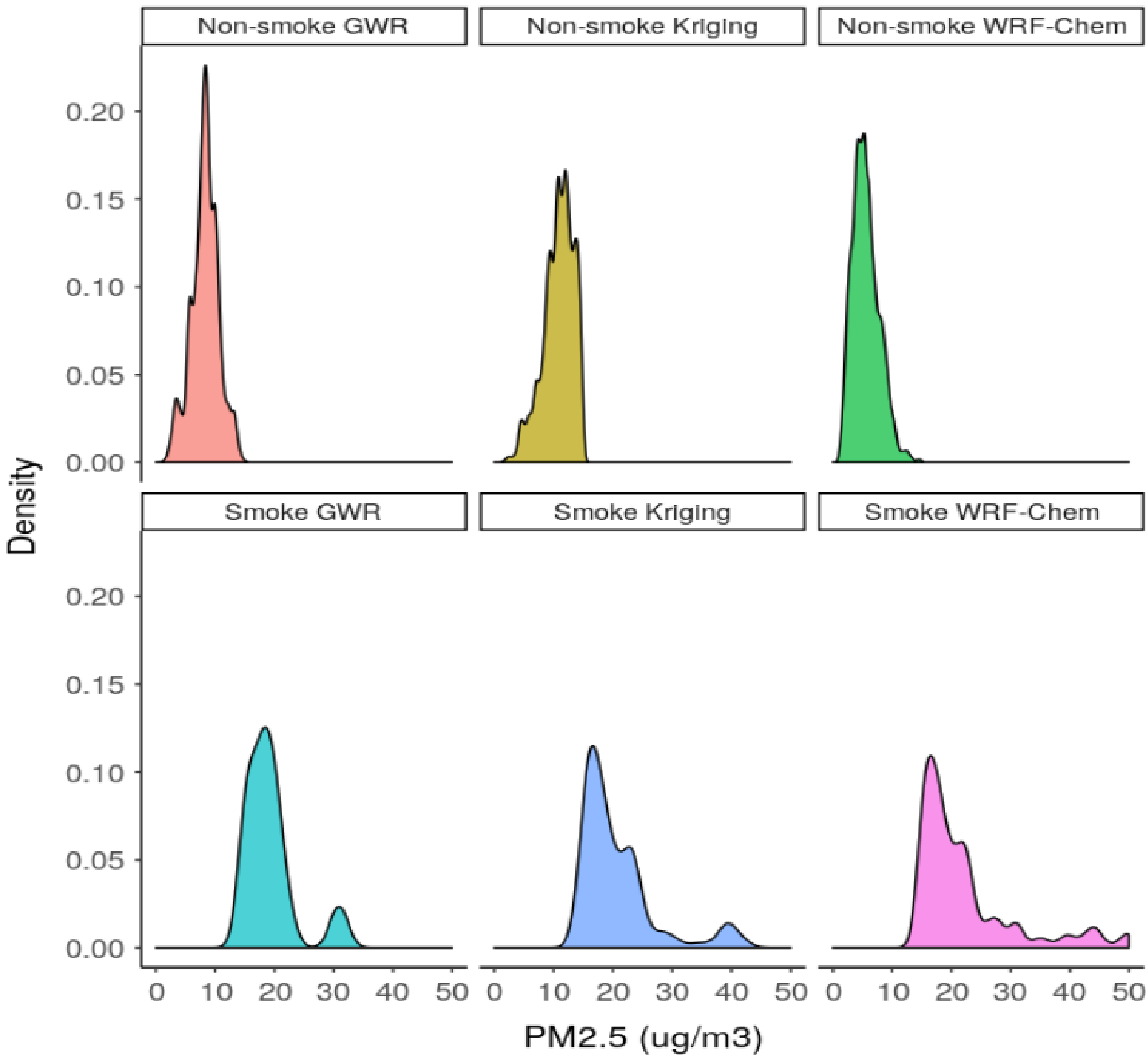


Figure A10. $PM_{2.5}$ distribution of 20 μg/m^3 threshold method wildfire smoke and non-wildfire smoke days, misclassified by Gan et al. method

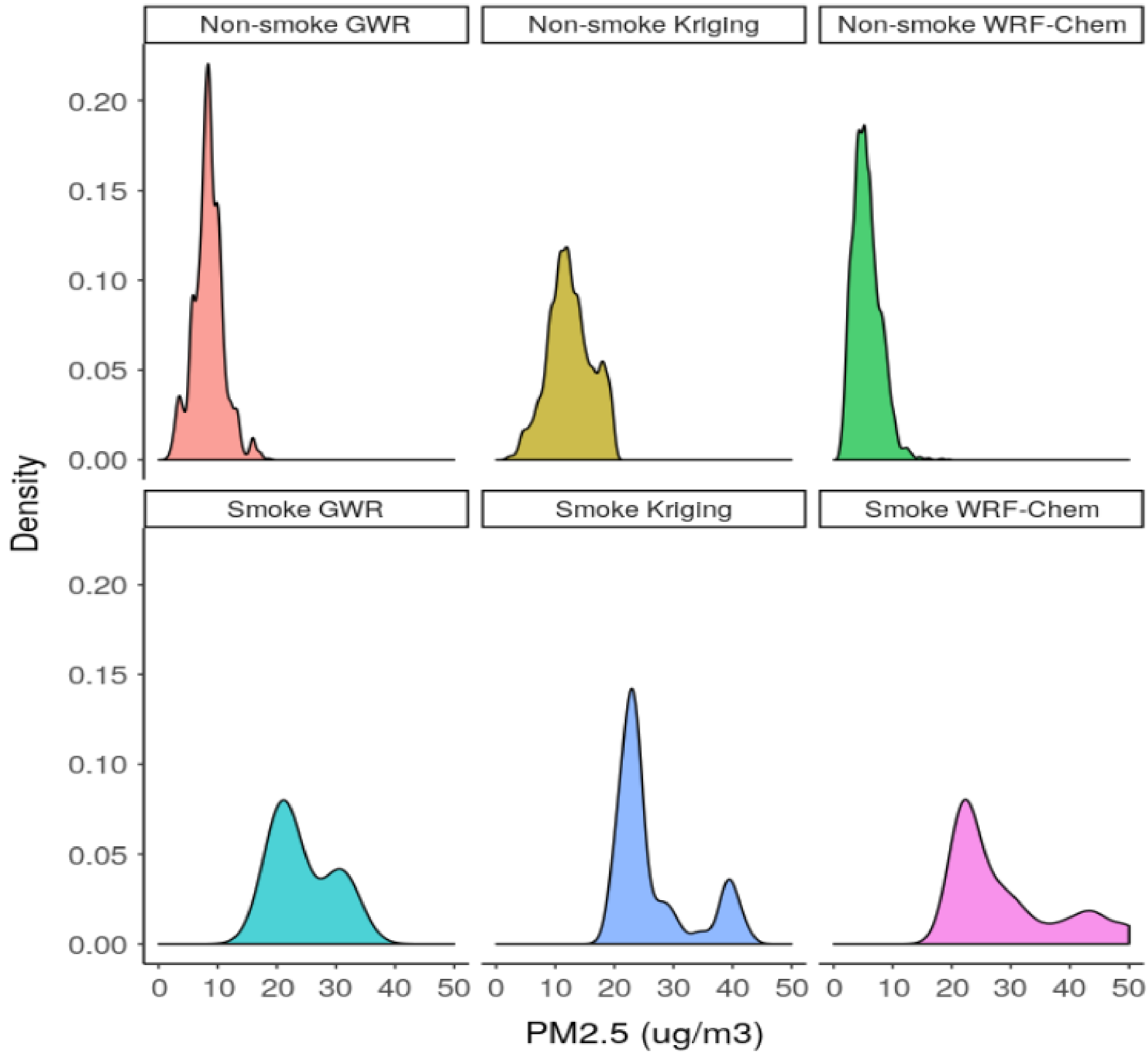
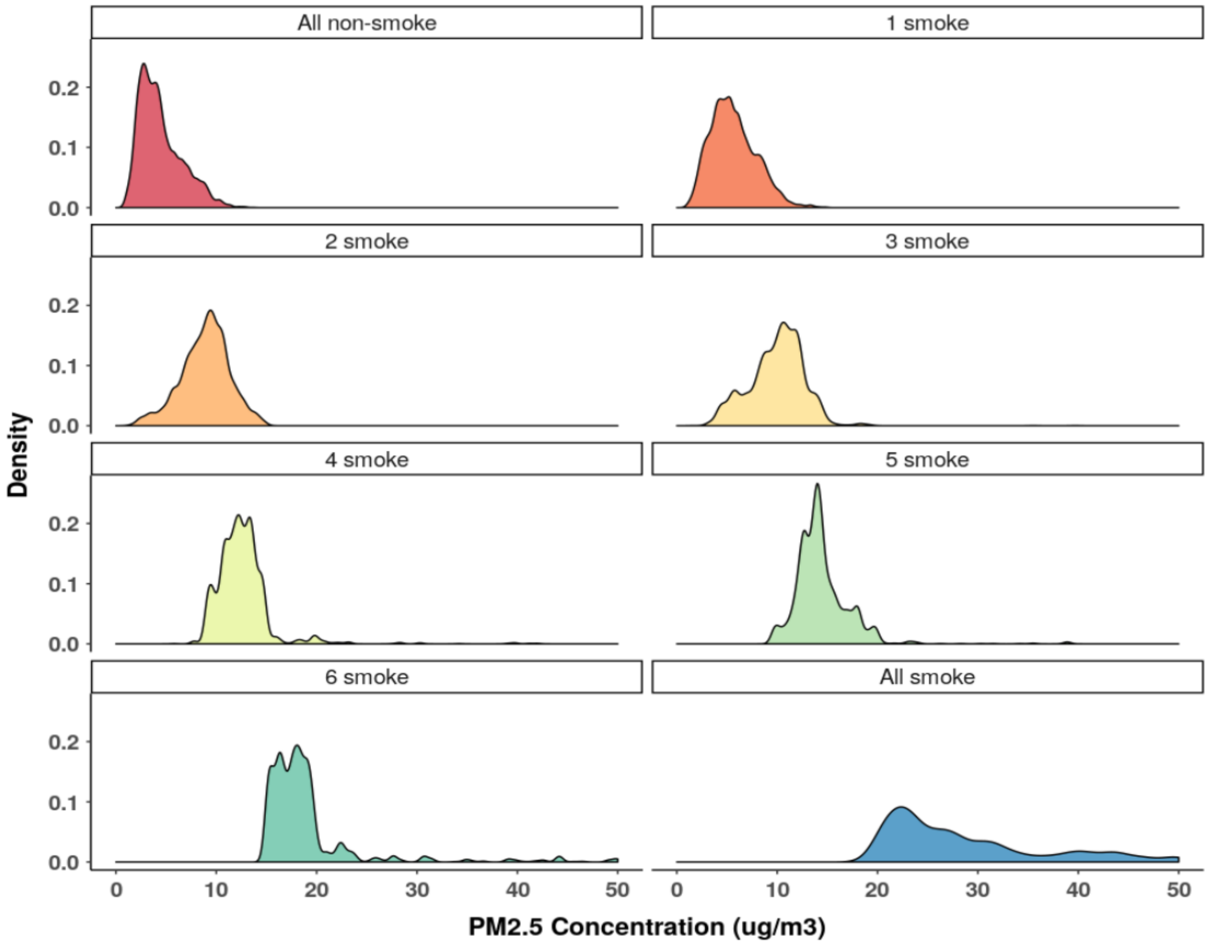


Figure A11. Distribution of $PM_{2.5}$ by number of definitions classifying days as smoke-affected[1]



[1]$PM_{2.5}$ distribution cut off at 50 $\mu g/m^3$ in order to visualize the distributions

Appendix B. Aim 2 additional plots and tables

Figure B1. Percent of summer days by WAQA category

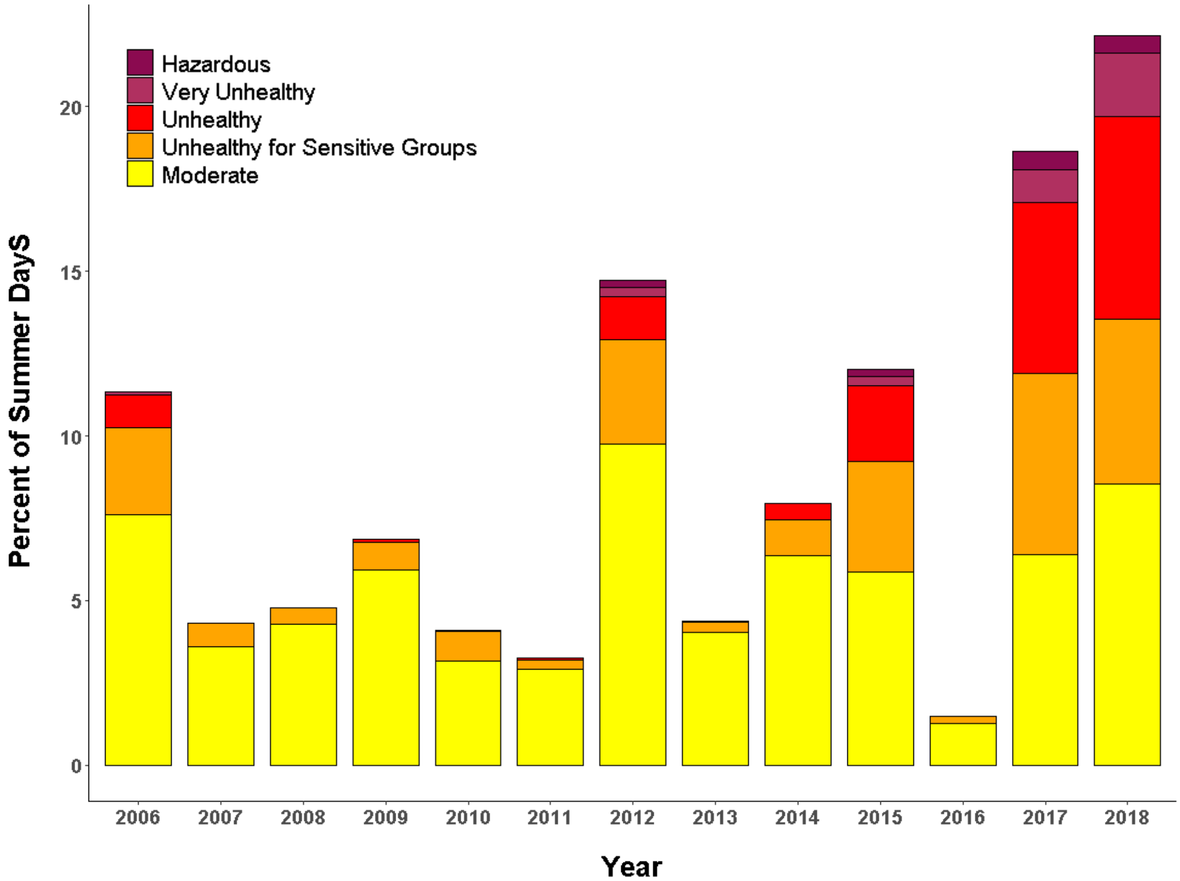


Figure B2. Boxplot of $PM_{2.5}$ on case and referent days for all deaths

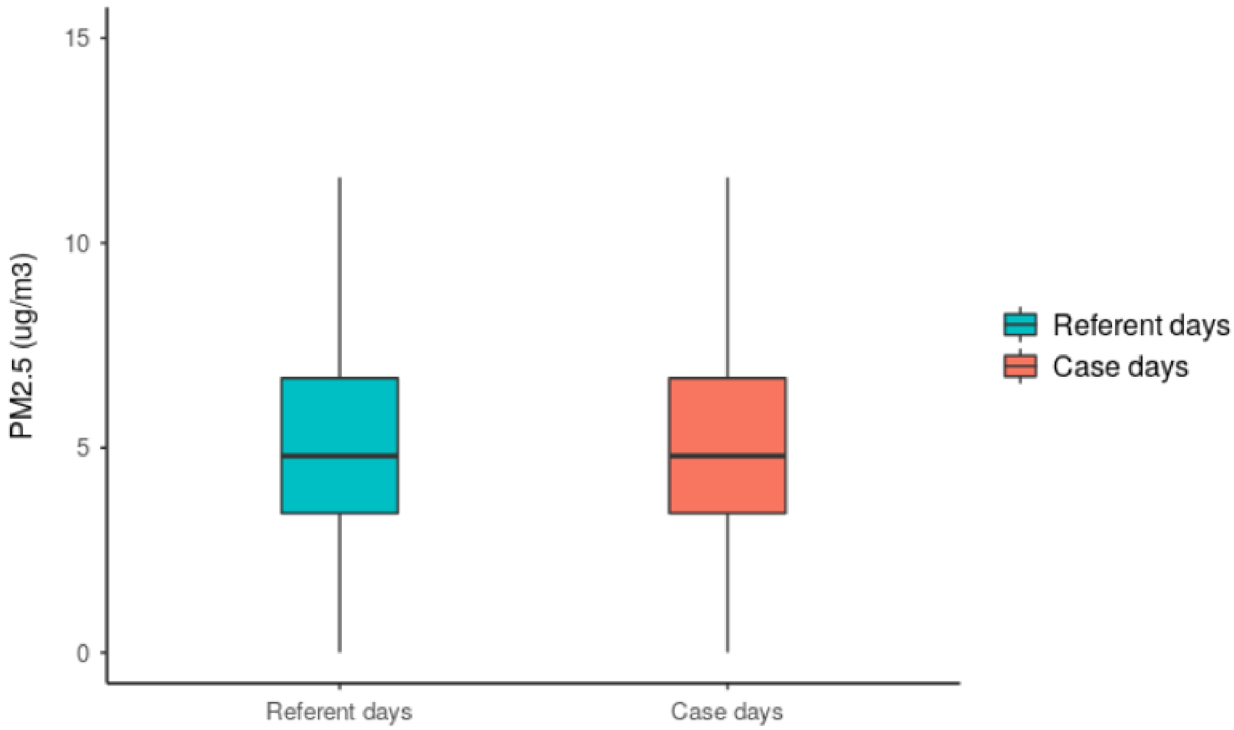


Figure B3. Boxplot of $PM_{2.5}$ on wildfire smoke days vs non-wildfire smoke days

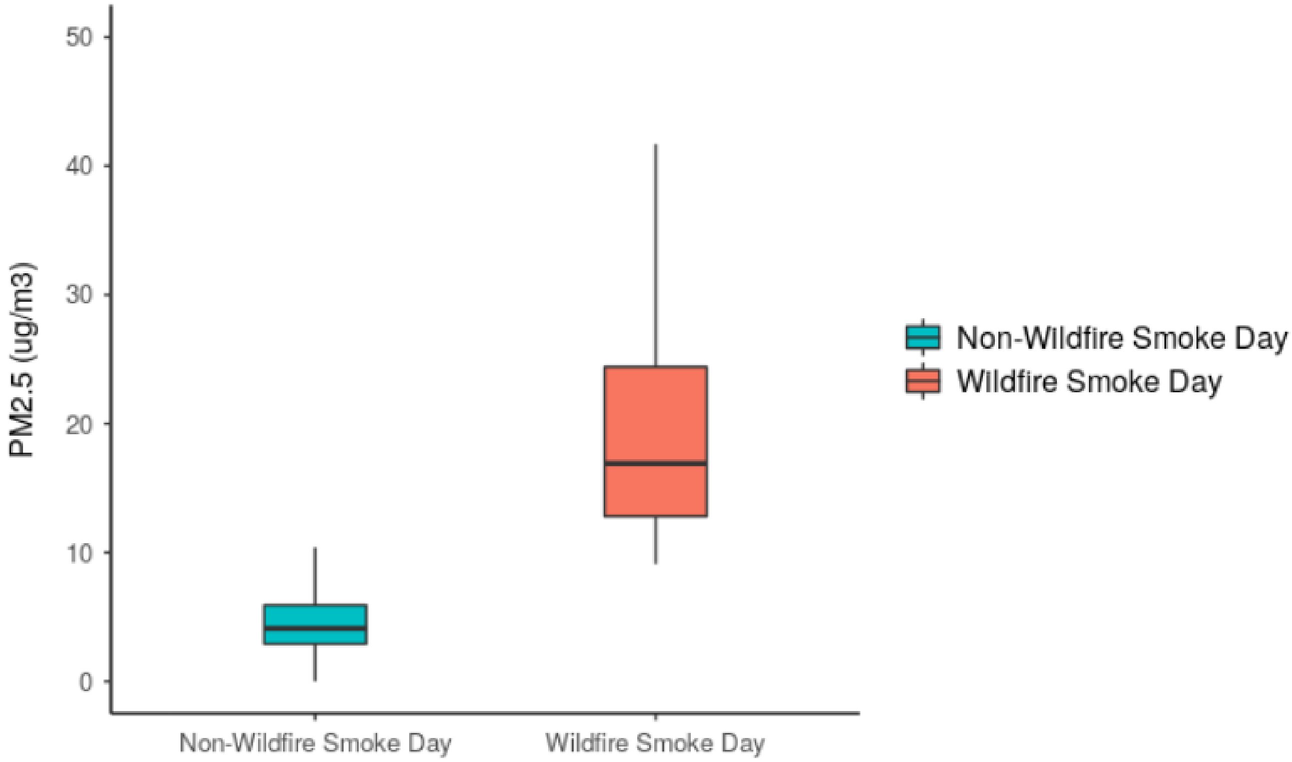


Figure B4. Boxplot of humidex on wildfire smoke days vs non-wildfire smoke days

Table B1. Results of a sensitivity analysis: ORs and 95% CIs for all-ages, all non-traumatic mortality associated with a >20.4 μg/m^3 threshold defining wildfire smoke days

Category	Adjusted OR (95% CI)	N (%)
All non-traumatic	1.00 (0.96, 1.04)	171,804 (100)
Age group		
0-4	2,301 (1.3)	2,301 (1.3)
5-14	668 (0.4)	668 (0.4)
15-44	4,969 (2.9)	4,969 (2.9)
45-64	32,375 (18.7)	32,375 (18.7)
65-84	75,110 (43.4)	75,110 (43.4)
85+	57,618 (33.3)	57,618 (33.3)
Cause of death		
Cardiovascular	0.96 (0.88, 1.03)	44,565 (25.9)
Ischemic heart disease	0.99 (0.82, 1.19)	7,945 (4.6)
Respiratory	1.10 (0.97, 1.24)	16,286 (9.5)
Asthma	0.56 (0.19, 1.71)	254 (0.14)
COPD	1.09 (0.93, 1.28)	9,571 (5.6)
Pneumonia	1.07 (0.73, 1.57)	2,174 (1.3)
Cerebrovascular	0.86 (0.66, 1.12)	3,747 (2.2)
Location		
Urban	1.01 (0.96, 1.06)	84,626 (48.9)
Non-urban	0.99 (0.93, 1.05)	88,415 (51.1)
Race		
White	1.00 (0.96, 1.04)	156,267 (90.3)
Black	0.99 (0.77, 1.27)	4,686 (2.7)
Native American	1.03 (0.76, 1.39)	2,449 (1.4)
Hispanic	0.74 (0.55, 1.00)	2,230 (1.3)
Native Hawaiian/ Other Pacific Islander	1.26 (0.88, 1.81)	1,635 (0.9)
Asian	1.18 (0.95, 1.47)	5,215 (3.0)
Median household income		
<$35,000	0.93 (0.82, 1.05)	16,092 (9.4)
$35,000 - $50,000	1.02 (0.94, 1.10)	41,919 (24.4)
$50,000 - $75,000	1.05 (0.99, 1.12)	63,056 (36.7)
$75,000 - $100,000	0.99 (0.90, 1.10)	24,631 (14.3)
≥$100,000	0.96 (0.81, 1.15)	8,171 (4.8)

* p ≤ 0.05, ** p ≤ 0.01, *** p ≤ 0.001

Appendix C. Methods

Data Cleaning and Preparation

This section describes data preparation and cleaning for the mortality and environmental data used in this project. The Washington State death statistical files and geocoded death files were purchased from the Washington State Department of Health for 2006-2017: https://www.doh.wa.gov/ DataandStatisticalReports/HealthStatistics/Death. The geocoded files were joined to create one file with all geocoded deaths for 2006-2017, and loaded into ArcGIS. The geocoded death file was joined with the 4x4km exposure grid (defined below) using the Spatial Join tool, and then output to yield a .csv file with each death assigned to a grid cell with a $PM_{2.5}$ and humidex value for the date of death. The output file was then merged on the death statistical files by the individual id. The final file contained the individual characteristics of each person who died during this period, and the environmental characteristics at the location of death.

The study exclusion criteria were then applied to the full mortality dataset: non-traumatic deaths in June-September of 2006-2017. For each individual characteristic of interest, proper stratification was applied to report counts, percent, and odds ratio for each category. An additional ecological variable was added to each individual death using American community survey's median household income by census tract. Median household income by census tract was available for 2010-2017. For 2006-2009, the average percent change at the census tract level between 2010-2016 was applied back to fill in estimates of median household income for the missing years across all census tracts. These were merged onto the mortality data file by census tract, county, and year.

To prepare the dataset for the time-stratified case-crossover analysis, each case had to be assigned referent days. The time-stratified design defines referent days as all days on the same day of the week and month of death. Each case was then appropriately assigned referent days, yielding a final dataset with multiple rows per individual.

Exposure Data Methods

Population exposure to $PM_{2.5}$ is based on Washington State University's AIRPACT $PM_{2.5}$ Model, which produces air quality forecasts for the Pacific Northwest (Washington State University, 2019). The AIRPACT model is based on a 4x4km grid and models air quality in the state based on emissions, meteorology, and pollutant chemistry and physics. The model predicts hourly $PM_{2.5}$ concentrations for each grid cell, producing an hourly spatial field of $PM_{2.5}$ across the state.

The daily mean $PM_{2.5}$ values generated from the AIRPACT model exhibit both random and systematic bias in different parts of the state, necessitating small corrections in order to produce a more accurate $PM_{2.5}$ grid. Each air quality monitoring site in the state was matched to the grid cell closest to it from the AIRPACT model. Next, the ratio between the summer mean 24-hour $PM_{2.5}$ value from the model and the summer mean 24-hour $PM_{2.5}$ value from the monitors was calculated. Each monitor was then assigned a ratio, and these ratios were interpolated across the state using Empirical Bayesian Kriging, which is a geostatistical interpolation method that minimizes error over other kriging methods by estimating different semivariograms across subsets of the input data (Pilz & Spöck, 2008). The interpolated ratio at each 4x4km grid cell was then multiplied by the model mean $PM_{2.5}$ to end up with the estimated summertime means across the state, at a 4x4km resolution.

The 4x4km $PM_{2.5}$ grid was then overlaid with the Washington State Office of Financial Management's (OFM) yearly population estimates at the census block group level (Management, 2017). In cases where census block group boundaries cross grid cells, the percent of the population's census block group attributed to each grid cell was determined by the area that fell within the grid cell. This method assumes populations are evenly distributed within each census block group, and that the population within each grid cell lives at the centroid of the grid cell.

Next, we determined the number of active monitors for each year from 2006 to 2017 based on whether concentration data are available on at least 50% of summer days (defined as June through September). Then, within each year, we identified the three nearest active monitoring sites to each grid cell. We then compared the summertime mean for that year at each of the three nearest neighbors to the grid cell summertime mean. If any of the summertime means from the three nearest monitors were more than 2 $\mu g/m^3$ different from the grid cell mean, the monitor was dropped as a valid option for that grid cell on that year. At the end of this process, within each year from 2006-2017, each grid cell had from one to three possible monitors identified.

Then, for each day at each grid cell, if the closest monitor is active and met the above criteria, the concentration from this monitor is assigned to the grid cell. If the closest monitor fails to fulfill these criteria, this is tried with the second closest monitor, and then the third closest monitor, in order to minimize loss of data. We end with a list containing the following: on each day, every grid cell has an associated concentration and meteorological variables from a neighboring monitoring site with the population attributed to that grid cell.

Piecewise Linear Spline Analysis by WAQA category

Additional analyses were conducted to examine the odds of non-traumatic mortality according to the Washington Air Quality Advisory (WAQA) categories (WA DOH; WA ECY, n.d.). The results of a piecewise linear spline model (Table C1) indicate no category has a significant increase in odds of mortality. The odds of non-traumatic mortality are nearly identical for the first three categories (Good, Moderate, and Unhealthy for Sensitive Groups), then increases for the next category (Unhealth), followed by a decrease (Very Unhealthy) and another increase for the last category (Hazardous). This model includes the effect of anthropogenic $PM_{2.5}$, especially in the lower two WAQA categories (Good and Moderate). The table shows the number and percent of grid cell days by category across the time series, with relatively few grid cell days in the severe categories.

Table C1. Results of piecewise linear model, by WAQA category

WAQA Category	Adjusted OR (95%CI)	Number of grid cell days (%)
Good	1.03 (0.99, 1.07)	12,957,185 (92.45)
Moderate	1.02 (0.97, 1.08)	628,038 (4.48)
Unhealthy for Sensitive Groups	1.02 (0.95, 1.10)	248,871 (1.78)
Unhealthy	1.09 (0.92, 1.28)	144,841 (1.03)
Very Unhealthy	0.87 (0.68, 1.11)	24,212 (0.17)
Hazardous	1.03 (0.26, 4.1)	10,242 (0.07)

This study applied a piecewise linear fit to better estimate and translate health risks associated with wildfire smoke using a familiar air quality categorization in Washington: the Washington Air Quality Advisory (WAQA) (WA DOH; WA ECY, n.d.). The process of designating knots at each of the WAQA $PM_{2.5}$ concentration categories resulted in the inclusion of both anthropogenic and wildfire smoke-related $PM_{2.5}$, with anthropogenic $PM_{2.5}$ predominantly in the “Good” and “Moderate” categories. The remaining categories largely reflect days with affected by wildfire smoke. The overall trend shows a nearly identical odds of non-traumatic mortality from “Good” through “Unhealthy for Sensitive Groups”, followed by unstable estimates for the last three categories, with no significant estimates. A possible explanation for the decrease in odds of non-traumatic mortality at "Very Unhealthy" levels is as wildfire smoke intensity increases, individuals may be more likely to take protective measures, by staying indoors and reducing outdoor activity, thereby decreasing their exposure and odds of associated non-traumatic mortality. However, once $PM_{2.5}$ reaches the "Hazardous" category, indoor air quality is likely to be affected, possibly increasing exposure and the odds of non-traumatic mortality. However, the number of grid cell events in the more severe WAQA categories is relatively low, and does not give us the power to detect an effect if there is one, as indicated by the wide confidence intervals. With increasingly worse wildfire seasons in Washington state, and more days falling into the higher WAQA categories, additional power will be present to detect an effect in the higher categories if one is present. This study was only able to investigate years through 2017. 2017 was Washington's first statewide wildfire smoke event lasting multiple days at unhealthy air concentrations. The summer of 2018 followed with a more severe statewide event. Additional research incorporating the 2018 wildfire season in Washington and beyond may provide additional power to our study and yield more informative results.

Classification of Monitors

The Washington Air Monitoring Network is made up of 68 air quality monitors capturing a variety of pollutants, including fine particulate matter ($PM_{2.5}$). In our determination of wildfire smoke days versus non-wildfire smoke day, half of the monitors in each of three urban areas had to be above 9 $\mu g/m^3$ to be considered a wildfire smoke day. The three urban areas are Seattle, Tacoma, and Spokane. A list of the monitors in each urban area or cluster, is included below. All other monitors in the network were considered to be not urban for the purposes of this analysis.

<u>Seattle</u>: Lake Forest Park, Bellevue, Seattle-Duwamish, Seattle-South Park, Kent, Lynnwood, Seattle-10th and Weller, Seattle-Beacon Hill

<u>Tacoma</u>: Tacoma-S 36th, Tacoma-L, Tacoma-Alexander, Puyallup

<u>Spokane</u>: Spokane-Broadway, Spokane-Colbert-Greenbluff, Spokane-Monroe, Spokane-Augusta, Spokane-College, Airway Heights, Liberty Lake

Case-crossover methods

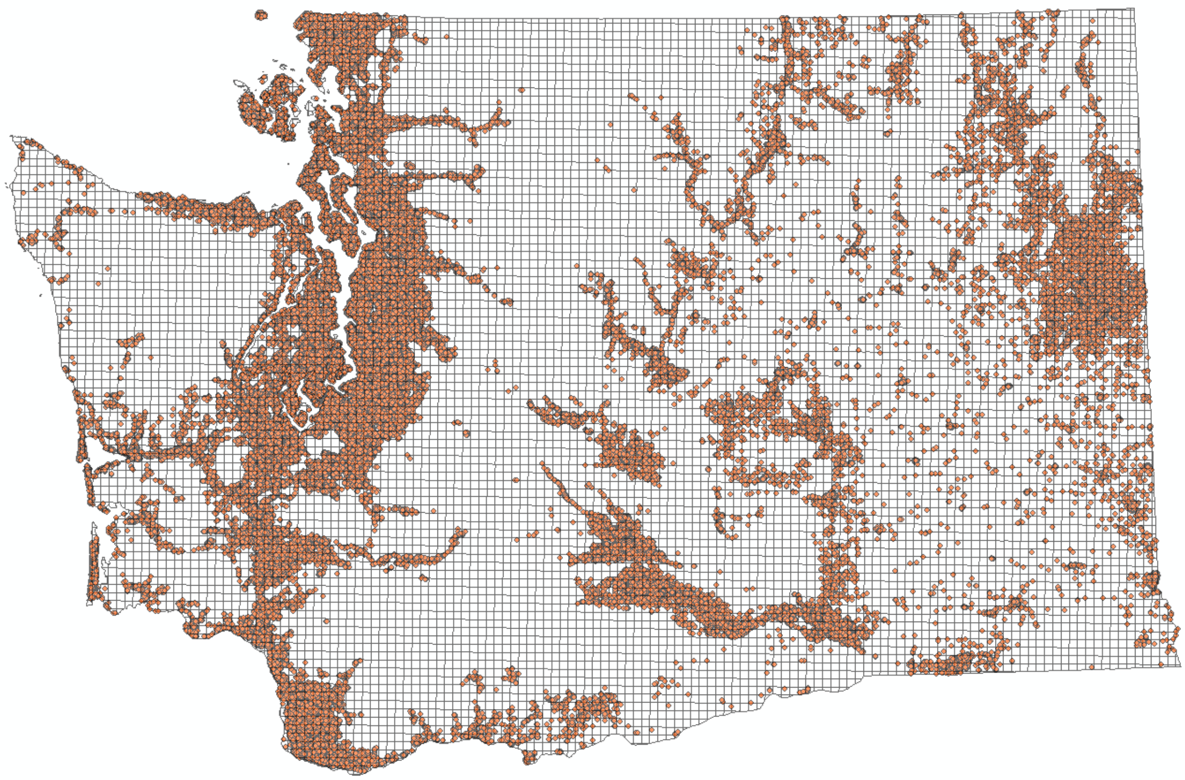
The case-crossover model is evaluated within strata, or within each individual, comparing the odds of death on wildfire smoke days compared to non-wildfire smoke days. Due to the nature of the design, only strata with a contrast of wildfire smoke days and non-wildfire smoke days contribute to the overall effect estimate. The model used for evaluating the odds of non-traumatic mortality on wildfire smoke days compared to non-wildfire smoke days is below, using conditional logistic regression:

$$\log(E(Y_t)) = \beta_0 + \beta_1 smoke.day + \beta_2 humidex + strata(id),$$

where $E(Y_t)$ is the expected count of non-traumatic deaths at time t, β_0 is the intercept, β_1 is the odds ratio and effect estimate of interest, β_2 controls for humidex, and $strata(id)$ indicates that the regression is evaluated within strata.

GIS Exposure Grid

Figure C1. Washington State 4x4 km exposure grid



Appendix D. Data Analysis Code

Program 1: Aim 1 data cleaning and preparation, part 1

```
```{r readin, echo=F}

spec.data <- fread("/home/gradstudent/doubleda/My Documents/MPH Year 2/Methods
paper/speciation_merged_with_pm25_data.csv")
clean.cutpoints <- fread("/home/gradstudent/doubleda/My Documents/MPH Year 2/Methods
paper/wildfire_smoke_cutoffs_by_site_and_county.csv")
```

```{r create.vars, echo=F}

#merge spec.data with clean.cutpoints by Site
wf.exp.df <- merge(spec.data, clean.cutpoints, by="Site")

wf.exp.select <- wf.exp.df %>% select(date, Site, OC_final, EC_Final, Conc_ugm3,
final_fire_flag, Max_PM2.5_noWF.site)

wf.exp.select$thresh.15 <- with(wf.exp.select, ifelse(Conc_ugm3 >= 15, 1, 0))
wf.exp.select$thresh.20 <- with(wf.exp.select, ifelse(Conc_ugm3 >= 20, 1, 0))
wf.exp.select$site.spec <- with(wf.exp.select, ifelse(Conc_ugm3 >= Max_PM2.5_noWF.site, 1,
0))
#wf.exp.select$year <- str_sub(wf.exp.select$date, -4, -1)

wf.exp.select$EC.OC.ratio <- with(wf.exp.select, EC_Final/OC_final)
wf.exp.select$date.fmt <- as.Date(wf.exp.select$date, "%m/%d/%Y")
```

```{r create.OC.flags, echo=F}

#low flag. all OC values below 1 ug/m3 should not be smoke days
wf.exp.select$OC.low.flag <- with(wf.exp.select, ifelse(OC_final < 1.667, 1, 0))
wf.exp.select$OC.high.flag <- with(wf.exp.select, ifelse(OC_final > 5, 1, 0))

#high flag. all OC values above 99% should be smoke days. overall.
#wf.exp.select$OC.per.95.all <- quantile(wf.exp.select$OC_final, c(.95))
#wf.exp.select$OC.per.99.all <- quantile(wf.exp.select$OC_final, c(.99))

#high flag by site. all OC values above 99% should be smoke days.
#wf.site.cutpoints <- wf.exp.select %>%
#  dplyr::group_by(Site) %>%
#  dplyr::summarise(OC.per.95.site = quantile(OC_final, c(.95)),
#                   OC.per.99.site = quantile(OC_final, c(.99)))

#wf.exp.update <- merge(wf.exp.select, wf.site.cutpoints, by="Site")
#wf.exp.update$OC.high.flag <- with(wf.exp.update, ifelse(OC_final > OC.per.95.site, 1, 0))
```
```

```
#set EC/OC ratio cutpoints.
#wf.exp.update$EC.OC.low.flag <- with(wf.exp.update, ifelse(EC.OC.ratio < 0.065, 1, 0))
#wf.exp.update$EC.OC.high.flag <- with(wf.exp.update, ifelse(EC.OC.ratio > 0.14, 1, 0))

#final df
wf.exp.final <- wf.exp.select %>% select(Site, date.fmt, Conc_ugm3, OC_final, final_fire_flag,
site.spec, thresh.15, thresh.20, OC.low.flag, OC.high.flag)
```

```{r misclass.calc, echo=F}

#calculate by site
#1) Percent of days where OC.low.flag = 1 and thresh.15 = 0 -- correctly classified (%)
#2) Percent of days where OC.high.flag = 1 and thresh.15 = 1 -- correclty classified (%)
#3) Percent of days where EC.OC.low.flag = 1 and thresh.15 = 0 - correctly classified
#4) Percent of days where EC.OC.high.flag = 1 and thresh.15 = 1 - correclty classified

#By site
wf.site.misclass <- wf.exp.final %>%
  dplyr::group_by(Site) %>%
  dplyr::summarise(num.days = n(),
            thresh.15.OC.low     = ((length(which(OC.low.flag==1 & thresh.15==1))/
                          length(which(OC.low.flag == 1)))*100),
            thresh.15.OC.high    = ((length(which(OC.high.flag==1 & thresh.15==0))/
                            length(which(OC.high.flag==1)))*100),
            thresh.20.OC.low     = ((length(which(OC.low.flag==1 & thresh.20==1))/
                            length(which(OC.low.flag == 1)))*100),
            thresh.20.OC.high    = ((length(which(OC.high.flag==1 & thresh.20==0))/
                            length(which(OC.high.flag==1)))*100),
            site.spec.OC.low     = ((length(which(OC.low.flag==1 & site.spec==1))/
                            length(which(OC.low.flag == 1)))*100),
            site.spec.OC.high    = ((length(which(OC.high.flag==1 & site.spec==0))/
                            length(which(OC.high.flag==1)))*100),
            final.flag.OC.low    = ((length(which(OC.low.flag==1 & final_fire_flag==1))/
                            length(which(OC.low.flag == 1)))*100),
            final.flag.OC.high   = ((length(which(OC.high.flag==1 & final_fire_flag==0))/
                            length(which(OC.high.flag==1)))*100)) %>%
  as.data.frame()

#Overall
wf.misclass <- wf.exp.final %>%
  dplyr::summarise(num.days = n(),
            thresh.15.OC.low     = ((length(which(OC.low.flag==1 & thresh.15==1))/
                          length(which(OC.low.flag == 1)))*100),
            thresh.15.OC.high    = ((length(which(OC.high.flag==1 & thresh.15==0))/
                            length(which(OC.high.flag==1)))*100),
```

```
            thresh.20.OC.low     = ((length(which(OC.low.flag==1 & thresh.20==1))/
                                   length(which(OC.low.flag == 1)))*100),
            thresh.20.OC.high    = ((length(which(OC.high.flag==1 & thresh.20==0))/
                                   length(which(OC.high.flag==1)))*100),
            site.spec.OC.low     = ((length(which(OC.low.flag==1 & site.spec==1))/
                                   length(which(OC.low.flag == 1)))*100),
            site.spec.OC.high    = ((length(which(OC.high.flag==1 & site.spec==0))/
                                   length(which(OC.high.flag==1)))*100),
            final.flag.OC.low    = ((length(which(OC.low.flag==1 & final_fire_flag==1))/
                                   length(which(OC.low.flag == 1)))*100),
            final.flag.OC.high   = ((length(which(OC.high.flag==1 & final_fire_flag==0))/
                                   length(which(OC.high.flag==1)))*100)) %>%
  as.data.frame()

wf.misclass.final <- cbind(Site = "All sites", wf.misclass)
wf.misclass.table <- rbind(wf.site.misclass, wf.misclass.final)
################################################################################
wf.num.days <- wf.exp.final %>%
  dplyr::group_by(Site) %>%
  dplyr::summarise(num.days = n(),
            per.days.low.OC = length(which(OC.low.flag==1))/num.days*100,
            per.days.high.OC = length(which(OC.high.flag==1))/num.days*100,
            per.thresh.15 = length(which(thresh.15==1))/num.days*100,
            per.thresh.20 = length(which(thresh.20==1))/num.days*100,
            per.site.spec = length(which(site.spec==1))/num.days*100,
            per.final.flag = length(which(final_fire_flag==1))/num.days*100) %>%
  as.data.frame()

wf.days.all <- wf.exp.final %>%
  dplyr::summarise(num.days = n(),
            per.days.low.OC = length(which(OC.low.flag==1))/num.days*100,
            per.days.high.OC = length(which(OC.high.flag==1))/num.days*100,
            per.thresh.15 = length(which(thresh.15==1))/num.days*100,
            per.thresh.20 = length(which(thresh.20==1))/num.days*100,
            per.site.spec = length(which(site.spec==1))/num.days*100,
            per.final.flag = length(which(final_fire_flag==1))/num.days*100) %>%
  as.data.frame()

wf.days.final <- cbind(Site = "All sites", wf.days.all)
wf.days.table <- rbind(wf.num.days, wf.days.final)

fwrite(wf.misclass.table, "WFS.misclass.site.def.csv")
fwrite(wf.days.table, "WFS.count.days.csv")
```

```{r exploratory, echo=F}
```

```
#Look at days where OC > 5 but no fire flag. first create that dataset.
high.OC.df <- with(wf.exp.select, subset(wf.exp.select, OC.high.flag == 1 &
                              final_fire_flag == 0))

high.OC.df.spec <- with(wf.exp.select, subset(wf.exp.select, OC.high.flag == 1 &
                                      site.spec == 0))

high.OC.df.thresh20 <- with(wf.exp.select, subset(wf.exp.select, OC.high.flag == 1 &
                                      thresh.20 == 0))

high.OC.df.thres15 <- with(wf.exp.select, subset(wf.exp.select, OC.high.flag == 1 &
                                      thresh.15 == 0))
```

```{r density.plot, echo=F, fig.width = 8, fig.height=6}

#create density plot of OC for final_fire_flag = 0 and 1. cut OC at 15 for display. howmany
dropped? 14

wf.exp.final$flag <- factor(wf.exp.final$final_fire_flag)

OC.compare <- ggplot(wf.exp.final, aes(x=OC_final, fill=flag)) +
                geom_density(alpha = 0.5) + xlim(0,15)
OC.compare

#by method. convert wide to long
wf.exp.long <- gather(wf.exp.final, method, flag, final_fire_flag:thresh.20, factor_key =F)
wf.exp.long$flag <- factor(wf.exp.long$flag)
wf.exp.long$method <- factor(wf.exp.long$method)
levels(wf.exp.long$method) <- c("Criteria method", "Site specific method", "> 15 ug/m3", "> 20
ug/m3")
levels(wf.exp.long$flag) <- c("Non-smoke day", "Smoke day")

OC.method.compare <- ggplot(wf.exp.long, aes(x=OC_final, fill=flag)) +
                    geom_density(alpha = 0.5) +
                    facet_wrap(~method) + xlim(0,15) +
                    xlab("Organic Carbon (ug/m3)") + ylab("Density") +
                    theme_bw() +
                    theme(panel.grid.major = element_blank(), panel.grid.minor =
element_blank()) +
                    theme(axis.title.y = element_text(margin = margin(r = 15), size = 12),
                    axis.title.x = element_text(margin = margin(t = 15), size = 12),
                    legend.title=element_blank(),
                    legend.text = element_text(margin = margin(l = 5), size = 10),
                    strip.text = element_text(size=12),
                    axis.text = element_text(size=10))
OC.method.compare
```

````
```
```{r misclass.all, echo=F}
#Overall
wf.mis.low.num <- wf.exp.final %>%
  dplyr::summarise(thresh.15.OC.low.WFS     = length(which(OC.low.flag==1 & thresh.15==1)),
                   thresh.15.OC.low.NS      = length(which(OC.low.flag==1 & thresh.15==0)),
                   thresh.20.OC.low.WFS     = length(which(OC.low.flag==1 & thresh.20==1)),
                   thresh.20.OC.low.NS      = length(which(OC.low.flag==1 & thresh.20==0)),
                   site.spec.OC.low.WFS     = length(which(OC.low.flag==1 & site.spec==1)),
                   site.spec.OC.low.NS      = length(which(OC.low.flag==1 & site.spec==0)),
                   final.flag.OC.low.WFS    = length(which(OC.low.flag==1 & final_fire_flag==1)),
                   final.flag.OC.low.NS     = length(which(OC.low.flag==1 & final_fire_flag==0)))
%>%
  as.data.frame()

wf.mis.low.per <- wf.exp.final %>%
  dplyr::summarise(thresh.15.OC.low.WFS     = ((length(which(OC.low.flag==1 &
thresh.15==1))/
                                        length(which(OC.low.flag == 1)))*100),
                   thresh.15.OC.low.NS     = ((length(which(OC.low.flag==1 & thresh.15==0))/
                                       length(which(OC.low.flag == 1)))*100),
                   thresh.20.OC.low.WFS     = ((length(which(OC.low.flag==1 & thresh.20==1))/
                                        length(which(OC.low.flag == 1)))*100),
                   thresh.20.OC.low.NS     = ((length(which(OC.low.flag==1 & thresh.20==0))/
                                        length(which(OC.low.flag == 1)))*100),
                   site.spec.OC.low.WFS     = ((length(which(OC.low.flag==1 & site.spec==1))/
                                        length(which(OC.low.flag == 1)))*100),
                   site.spec.OC.low.NS     = ((length(which(OC.low.flag==1 & site.spec==0))/
                                        length(which(OC.low.flag == 1)))*100),
                   final.flag.OC.low.WFS    = ((length(which(OC.low.flag==1 & final_fire_flag==1))/
                                        length(which(OC.low.flag == 1)))*100),
                   final.flag.OC.low.NS    = ((length(which(OC.low.flag==1 & final_fire_flag==0))/
                                        length(which(OC.low.flag == 1)))*100)) %>%
  as.data.frame()

#create table
wf.mis.low.numt <- t(wf.mis.low.num)
wf.mis.low.pert <- t(wf.mis.low.per)

wf.mis.low <- cbind(wf.mis.low.numt, wf.mis.low.pert)

wf.mis.hi.num <- wf.exp.final %>%
  dplyr::summarise(thresh.15.OC.high.WFS    = length(which(OC.high.flag==1 &
thresh.15==1)),
                   thresh.15.OC.high.NS    = length(which(OC.high.flag==1 & thresh.15==0)),
                   thresh.20.OC.high.WFS    = length(which(OC.high.flag==1 & thresh.20==1)),
````

```
            thresh.20.OC.high.NS    = length(which(OC.high.flag==1 & thresh.20==0)),
            site.spec.OC.high.WFS    = length(which(OC.high.flag==1 & site.spec==1)),
            site.spec.OC.high.NS    = length(which(OC.high.flag==1 & site.spec==0)),
            final.flag.OC.high.WFS   = length(which(OC.high.flag==1 & final_fire_flag==1)),
            final.flag.OC.high.NS   = length(which(OC.high.flag==1 & final_fire_flag==0)))
%>%
  as.data.frame()

wf.mis.hi.per <- wf.exp.final %>%
  dplyr::summarise(thresh.15.OC.high.WFS    = ((length(which(OC.high.flag==1 &
thresh.15==1))/
                              length(which(OC.high.flag==1)))*100),
            thresh.15.OC.high.NS    = ((length(which(OC.high.flag==1 & thresh.15==0))/
                              length(which(OC.high.flag==1)))*100),
            thresh.20.OC.high.WFS    = ((length(which(OC.high.flag==1 & thresh.20==1))/
                              length(which(OC.high.flag==1)))*100),
            thresh.20.OC.high.NS    = ((length(which(OC.high.flag==1 & thresh.20==0))/
                              length(which(OC.high.flag==1)))*100),
            site.spec.OC.high.WFS    = ((length(which(OC.high.flag==1 & site.spec==1))/
                              length(which(OC.high.flag==1)))*100),
            site.spec.OC.high.NS    = ((length(which(OC.high.flag==1 & site.spec==0))/
                              length(which(OC.high.flag==1)))*100),
            final.flag.OC.high.WFS   = ((length(which(OC.high.flag==1 & final_fire_flag==1))/
                              length(which(OC.high.flag==1)))*100),
            final.flag.OC.high.NS   = ((length(which(OC.high.flag==1 & final_fire_flag==0))/
                              length(which(OC.high.flag==1)))*100)) %>%
  as.data.frame()

#create table
wf.mis.hi.numt <- t(wf.mis.hi.num)
wf.mis.hi.pert <- t(wf.mis.hi.per)

wf.mis.hi <- cbind(wf.mis.hi.numt, wf.mis.hi.pert)

wf.mis.final <- cbind(wf.mis.low, wf.mis.hi)
kable(wf.mis.final, col.names=c("#", "%", "#", "%"), digits= 2) %>%
  kable_styling(bootstrap_options = c("striped")) %>%
  add_header_above(c("Method" = 1, "Days < 1.667 ug/m3 OC" = 2, "Days > 5 ug/m3 OC" =2))
```

Program 2: Aim 1 data cleaning and preparation, part 2

```
```{r readin, echo=F}
gan.krig <- fread("/home/gradstudent/doubleda/My Documents/MPH Year 2/Methods paper/kriging_pm25_wash2012.csv")
gan.gwr <- fread("/home/gradstudent/doubleda/My Documents/MPH Year 2/Methods paper/geo_weighted_ridge_regression_pm25_wash2012.csv")
gan.wrf <- fread("/home/gradstudent/doubleda/My Documents/MPH Year 2/Methods paper/wrf_chem_pm25_wash2012.csv")
gan.background <- fread("/home/gradstudent/doubleda/My Documents/MPH Year 2/Methods paper/background_pm25_wash2012.csv")

gan.gis <- gan.krig %>% select(WRFGRID_ID, Longitude, Latitude)
fwrite(gan.gis, "/home/gradstudent/doubleda/My Documents/MPH Year 2/Methods paper/gan.gis.csv")
clean.cutpoints <- fread("/home/gradstudent/doubleda/My Documents/MPH Year 2/Methods paper/wildfire_smoke_cutoffs_by_site_and_county.csv")
id.monitor.xwalk <- fread("/home/gradstudent/doubleda/My Documents/MPH Year 2/Methods paper/AQS.ID_Monitor_xwalk.csv")
```
```{r wide.to.long, echo=F}
#Convert all 3 from wide to long, assign fire flags. merge them on each other so have one df with id, lat/long, date, concentrations for each method and flags for each method. Export. Spatial join in GIS with grids. Then compare flags to these

#-----------------------------------Krigged dataset------------------------------------------
#convert ID column to factor
gan.krig$WRFGRID_ID <- factor(gan.krig$WRFGRID_ID)
gan.krig.long <- gather(gan.krig, date, krig.conc, `20120701`:`20121031`, factor_key=TRUE)
gan.krig.long$krig.flag <- with(gan.krig.long, ifelse(krig.conc > 10, 1, 0))

#-----------------------------------Geo-weighted ridge regression-----------------------------
gan.gwr$WRFGRID_ID <- factor(gan.gwr$WRFGRID_ID)
gan.gwr.long <- gather(gan.gwr, date, gwr.conc, `20120701`:`20121031`, factor_key=TRUE)
gan.gwr.long$gwr.flag <- with(gan.gwr.long, ifelse(gwr.conc > 10, 1, 0))

#-----------------------------------WRF-Chem----------------------------------
gan.wrf$WRFGRID_ID <- factor(gan.wrf$WRFGRID_ID)
gan.wrf.long <- gather(gan.wrf, date, wrf.conc, `20120701`:`20121031`, factor_key=TRUE)
gan.wrf.long$wrf.flag <- with(gan.wrf.long, ifelse(wrf.conc > 10, 1, 0))

#-----------------------------------Merge and export-----------------------------------
gan.krig.gwr <- merge(gan.krig.long, gan.gwr.long, by=c("WRFGRID_ID", "Longitude", "Latitude", "date"))
gan.all.methods <- merge(gan.krig.gwr, gan.wrf.long, by=c("WRFGRID_ID", "Longitude", "Latitude", "date"))
```
```

```
fwrite(gan.all.methods, "/home/gradstudent/doubleda/My Documents/MPH Year 2/Methods
paper/gan.all.methods.csv")
```
```{r load.data, echo=F}
#readin joined df. merge on gan.all.methods
gan.gis.join <- fread("/home/gradstudent/doubleda/My Documents/MPH Year 2/Methods
paper/gan_grid_join.csv")
exposure.grid <- fread("/home/gradstudent/doubleda/My Documents/MPH Year 2/Methods
paper/Population_by_grid_cells_w_fire_flags_FINAL.csv")

gan.all.methods$WRFGRID_ID <- as.numeric(as.character(gan.all.methods$WRFGRID_ID))
gan.gis.join.subset <- subset(gan.gis.join, gan.gis.join$Join_Count==1)
gan.gis.select <- gan.gis.join.subset %>% select(WRFGRID_ID, FIRST_xcel)
gan.merge <- merge(gan.gis.select, gan.all.methods, by=c("WRFGRID_ID"))
setnames(gan.merge, old=c("FIRST_xcel"), new=c("xcell_ycell"))
gan.merge$date <- as.character(gan.merge$date)

monitor.xwalk <- id.monitor.xwalk %>% select(xcell_ycell, Site, date)
exp.grid.merge <- merge(exposure.grid, monitor.xwalk, by=c("xcell_ycell", "date"))
exp.grid.select <- exp.grid.merge %>% select(xcell_ycell, date, Conc_ugm3, final_fire_flag,
Site)
exp.grid.select$date <- gsub("-", "", exp.grid.select$date)
merge.gan.exp <- merge(gan.merge, exp.grid.select, by=c("xcell_ycell", "date"))
gan.exp.final <- merge(merge.gan.exp, clean.cutpoints, by="Site")
gan.exp.final$thresh.15 <- with(gan.exp.final, ifelse(Conc_ugm3 >= 15, 1, 0))
gan.exp.final$thresh.20 <- with(gan.exp.final, ifelse(Conc_ugm3 >= 20, 1, 0))
gan.exp.final$site.spec <- with(gan.exp.final, ifelse(Conc_ugm3 >= Max_PM2.5_noWF.site, 1,
0))
```
```{r misclass, echo=F}
gan.exp.final$krig.flag2 <- ifelse(gan.exp.final$krig.flag == 0, "Non-smoke day",
                    ifelse(gan.exp.final$krig.flag == 1, "Smoke day", NA))
gan.exp.final$gwr.flag2 <- ifelse(gan.exp.final$gwr.flag == 0, "Non-smoke day",
                    ifelse(gan.exp.final$gwr.flag == 1, "Smoke day", NA))
gan.exp.final$wrf.flag2 <- ifelse(gan.exp.final$wrf.flag == 0, "Non-smoke day",
                    ifelse(gan.exp.final$wrf.flag == 1, "Smoke day", NA))

#for each gan method, compare flags.
#-------------------criteria method-------------------------
#with(gan.exp.final, table(final_fire_flag, krig.flag))
gan.exp.final$crit.krig.sm.miss <- ifelse(gan.exp.final$final_fire_flag==1 &
                              gan.exp.final$krig.flag=="Non-smoke day", 1, 0)
df.crit.krig.sm.miss <- subset(gan.exp.final, gan.exp.final$crit.krig.sm.miss==1)
#hist(df.crit.krig.sm.miss$Conc_ugm3)

gan.exp.final$crit.krig.nonsm.miss <- ifelse(gan.exp.final$final_fire_flag==0 &
```

```
                                     gan.exp.final$krig.flag=="Smoke day", 1, 0)
df.crit.krig.nonsm.miss <- subset(gan.exp.final, gan.exp.final$crit.krig.nonsm.miss==1)
#hist(df.crit.krig.nonsm.miss$Conc_ugm3)

#with(gan.exp.final, table(final_fire_flag, gwr.flag))
with(gan.exp.final, table(final_fire_flag, wrf.flag))

#-------------------site specific method-------------------------
#with(gan.exp.final, table(site.spec, krig.flag))
#with(gan.exp.final, table(site.spec, gwr.flag))
with(gan.exp.final, table(site.spec, wrf.flag))
#-------------------thresh 20 ug/m3-------------------------
#with(gan.exp.final, table(thresh.20, krig.flag))
#with(gan.exp.final, table(thresh.20, gwr.flag))
with(gan.exp.final, table(thresh.20, wrf.flag))
#-------------------thresh 15 ug/m3-------------------------
#with(gan.exp.final, table(thresh.15, krig.flag))
#with(gan.exp.final, table(thresh.15, gwr.flag))
with(gan.exp.final, table(thresh.15, wrf.flag))
#-------------------criteria method misclassification-------------------------
#create variable for criteria misclass type
gan.exp.final$crit.misclass <- with(gan.exp.final,
                      ifelse(final_fire_flag == 1 & krig.flag2=="Non-smoke day", "Smoke Kriging",
                      ifelse(final_fire_flag == 0 & krig.flag2 == "Smoke day", "Non-smoke Kriging",
                      ifelse(final_fire_flag == 1 & gwr.flag2 == "Non-smoke day", "Smoke GWR",
                      ifelse(final_fire_flag == 0 & gwr.flag2 == "Smoke day", "Non-smoke GWR",
                      ifelse(final_fire_flag == 1 & wrf.flag2 == "Non-smoke day", "Smoke WRF-
Chem",
                      ifelse(final_fire_flag == 0 & wrf.flag2 =="Smoke day", "Non-smoke WRF-
Chem", NA)))))))

#create faceted histogram for the criteria method
criteria.misclass <- ggplot(data=subset(gan.exp.final, !is.na(crit.misclass)), aes(Conc_ugm3)) +
  geom_histogram(aes(y = ..density.., fill=crit.misclass)) +
  facet_wrap(~crit.misclass) + xlim(0,50) +
  labs(x="PM2.5 (ug/m3)", y="Density", fill="Misclassification Type",
       title = "Criteria misclassification") +
  theme_classic() +
  theme(axis.title.y = element_text(margin = margin(r = 10), size = 12),
        axis.text.x = element_text(margin=margin(t=5), size = 10),
        axis.text.y = element_text(margin=margin(r=5), size = 10))
criteria.misclass

#-------------------site-specific method misclassification-------------------------
#create variable for site specific misclass type
gan.exp.final$site.misclass <- with(gan.exp.final,
```

```
                     ifelse(site.spec == 1 & krig.flag2=="Non-smoke day", "Smoke Kriging",
                     ifelse(site.spec == 0 & krig.flag2 == "Smoke day", "Non-smoke Kriging",
                     ifelse(site.spec == 1 & gwr.flag2 == "Non-smoke day", "Smoke GWR",
                     ifelse(site.spec == 0 & gwr.flag2 == "Smoke day", "Non-smoke GWR",
                     ifelse(site.spec == 1 & wrf.flag2 == "Non-smoke day", "Smoke WRF-Chem",
                     ifelse(site.spec == 0 & wrf.flag2 =="Smoke day", "Non-smoke WRF-Chem",
NA)))))))

#create faceted histogram for the site specific method
site.misclass <- ggplot(data=subset(gan.exp.final, !is.na(site.misclass)), aes(Conc_ugm3)) +
  geom_histogram(aes(y = ..density.., fill=site.misclass)) +
  facet_wrap(~site.misclass) + xlim(0,50) +
  labs(x="PM2.5 (ug/m3)", y="Density", fill="Misclassification Type",
       title="Site-specific misclassification") +
  theme_classic() +
  theme(axis.title.y = element_text(margin = margin(r = 10), size = 12),
        axis.text.x = element_text(margin=margin(t=5), size = 10),
        axis.text.y = element_text(margin=margin(r=5), size = 10))
site.misclass

#-------------------20 ug/m3 threshold method misclassification--------------------------
#create variable for site specific misclass type
gan.exp.final$thr20.misclass <- with(gan.exp.final,
                     ifelse(thresh.20 == 1 & krig.flag2=="Non-smoke day", "Smoke Kriging",
                     ifelse(thresh.20 == 0 & krig.flag2 == "Smoke day", "Non-smoke Kriging",
                     ifelse(thresh.20 == 1 & gwr.flag2 == "Non-smoke day", "Smoke GWR",
                     ifelse(thresh.20 == 0 & gwr.flag2 == "Smoke day", "Non-smoke GWR",
                     ifelse(thresh.20 == 1 & wrf.flag2 == "Non-smoke day", "Smoke WRF-Chem",
                     ifelse(thresh.20 == 0 & wrf.flag2 =="Smoke day", "Non-smoke WRF-Chem",
NA)))))))

#create faceted histogram for the site specific method
thresh20.misclass <- ggplot(data=subset(gan.exp.final, !is.na(thr20.misclass)), aes(Conc_ugm3))
+ geom_histogram(aes(y = ..density.., fill=thr20.misclass)) +
  facet_wrap(~thr20.misclass) + xlim(0,50) +
  labs(x="PM2.5 (ug/m3)", y="Density", fill="Misclassification Type",
       title = "20 ug/m3 threshold misclassification") +
  theme_classic() +
  theme(axis.title.y = element_text(margin = margin(r = 10), size = 12),
        axis.text.x = element_text(margin=margin(t=5), size = 10),
        axis.text.y = element_text(margin=margin(r=5), size = 10))
thresh20.misclass

#-------------------15 ug/m3 threshold method misclassification--------------------------
#create variable for site specific misclass type
gan.exp.final$thr15.misclass <- with(gan.exp.final,
```

```
                          ifelse(thresh.15 == 1 & krig.flag2=="Non-smoke day", "Smoke Kriging",
                          ifelse(thresh.15 == 0 & krig.flag2 == "Smoke day", "Non-smoke Kriging",
                          ifelse(thresh.15 == 1 & gwr.flag2 == "Non-smoke day", "Smoke GWR",
                          ifelse(thresh.15 == 0 & gwr.flag2 == "Smoke day", "Non-smoke GWR",
                          ifelse(thresh.15 == 1 & wrf.flag2 == "Non-smoke day", "Smoke WRF-Chem",
                          ifelse(thresh.15 == 0 & wrf.flag2 =="Smoke day", "Non-smoke WRF-Chem",
NA)))))))

#create faceted histogram for the site specific method
thresh15.misclass <- ggplot(data=subset(gan.exp.final, !is.na(thr15.misclass)), aes(Conc_ugm3))
+ geom_histogram(aes(y = ..density.., fill=thr15.misclass)) +
  facet_wrap(~thr15.misclass) + xlim(0,50) +
  labs(x="PM2.5 (ug/m3)", y="Density", fill="Misclassification Type",
       title = " 15 ug/m3 threshold misclassification") +
  theme_classic() +
  theme(axis.title.y = element_text(margin = margin(r = 10), size = 12),
        axis.text.x = element_text(margin=margin(t=5), size = 10),
        axis.text.y = element_text(margin=margin(r=5), size = 10))
thresh15.misclass
```

```{r calc.mis, echo=F}
gan.compare.krig.num <- gan.exp.final %>%
  dplyr::summarise(crit.non.smoke = (length(which(final_fire_flag == 0 & krig.flag ==0))),
                   crit.smoke     = (length(which(final_fire_flag == 1 & krig.flag ==1))),
                   site.non.smoke = (length(which(site.spec == 0 & krig.flag ==0))),
                   site.smoke     = (length(which(site.spec == 1 & krig.flag ==1))),
                   thr20.non.smoke= (length(which(thresh.20 == 0 & krig.flag ==0))),
                   thr20.smoke    = (length(which(thresh.20 == 1 & krig.flag ==1))),
                   thr15.non.smoke= (length(which(thresh.15 == 0 & krig.flag ==0))),
                   thr15.smoke    = (length(which(thresh.15 == 1 & krig.flag ==1))))
gan.compare.krig <- gan.exp.final %>%
  dplyr::summarise(crit.non.smoke = (length(which(final_fire_flag == 0 & krig.flag ==0)))/
                                  length(which(final_fire_flag ==0))*100,
                   crit.smoke     = (length(which(final_fire_flag == 1 & krig.flag ==1)))/
                                  length(which(final_fire_flag ==1))*100,
                   site.non.smoke = (length(which(site.spec == 0 & krig.flag ==0)))/
                                  length(which(site.spec ==0))*100,
                   site.smoke     = (length(which(site.spec == 1 & krig.flag ==1)))/
                                  length(which(site.spec ==1))*100,
                   thr20.non.smoke= (length(which(thresh.20 == 0 & krig.flag ==0)))/
                                  length(which(thresh.20 ==0))*100,
                   thr20.smoke    = (length(which(thresh.20 == 1 & krig.flag ==1)))/
                                  length(which(thresh.20 ==1))*100,
                   thr15.non.smoke= (length(which(thresh.15 == 0 & krig.flag ==0)))/
                                  length(which(thresh.15 ==0))*100,
                   thr15.smoke    = (length(which(thresh.15 == 1 & krig.flag ==1)))/
```

```
                                      length(which(thresh.15 ==1))*100)
gan.compare.gwr <- gan.exp.final %>%
  dplyr::summarise(crit.non.smoke = (length(which(final_fire_flag == 0 & gwr.flag ==0)))/
                                      length(which(final_fire_flag ==0))*100,
                   crit.smoke     = (length(which(final_fire_flag == 1 & gwr.flag ==1)))/
                                      length(which(final_fire_flag ==1))*100,
                   site.non.smoke = (length(which(site.spec == 0 & gwr.flag ==0)))/
                                      length(which(site.spec ==0))*100,
                   site.smoke     = (length(which(site.spec == 1 & gwr.flag ==1)))/
                                      length(which(site.spec ==1))*100,
                   thr20.non.smoke= (length(which(thresh.20 == 0 & gwr.flag ==0)))/
                                      length(which(thresh.20 ==0))*100,
                   thr20.smoke    = (length(which(thresh.20 == 1 & gwr.flag ==1)))/
                                      length(which(thresh.20 ==1))*100,
                   thr15.non.smoke= (length(which(thresh.15 == 0 & gwr.flag ==0)))/
                                      length(which(thresh.15 ==0))*100,
                   thr15.smoke    = (length(which(thresh.15 == 1 & gwr.flag ==1)))/
                                      length(which(thresh.15 ==1))*100)
gan.compare.wrf <- gan.exp.final %>%
  dplyr::summarise(crit.non.smoke = (length(which(final_fire_flag == 0 & wrf.flag ==0)))/
                                      length(which(final_fire_flag ==0))*100,
                   crit.smoke     = (length(which(final_fire_flag == 1 & wrf.flag ==1)))/
                                      length(which(final_fire_flag ==1))*100,
                   site.non.smoke = (length(which(site.spec == 0 & wrf.flag ==0)))/
                                      length(which(site.spec ==0))*100,
                   site.smoke     = (length(which(site.spec == 1 & wrf.flag ==1)))/
                                      length(which(site.spec ==1))*100,
                   thr20.non.smoke= (length(which(thresh.20 == 0 & wrf.flag ==0)))/
                                      length(which(thresh.20 ==0))*100,
                   thr20.smoke    = (length(which(thresh.20 == 1 & wrf.flag ==1)))/
                                      length(which(thresh.20 ==1))*100,
                   thr15.non.smoke= (length(which(thresh.15 == 0 & wrf.flag ==0)))/
                                      length(which(thresh.15 ==0))*100,
                   thr15.smoke    = (length(which(thresh.15 == 1 & wrf.flag ==1)))/
                                      length(which(thresh.15 ==1))*100)
```
```{r calc.mis.2, echo=F}
gan.compare.krig2 <- gan.exp.final %>%
  dplyr::summarise(crit.non.smoke = (length(which(final_fire_flag == 0 & krig.flag ==0)))/
                                      length(which(krig.flag ==0))*100,
                   crit.smoke     = (length(which(final_fire_flag == 1 & krig.flag ==1)))/
                                      length(which(krig.flag ==1))*100,
                   site.non.smoke = (length(which(site.spec == 0 & krig.flag ==0)))/
                                      length(which(krig.flag ==0))*100,
                   site.smoke     = (length(which(site.spec == 1 & krig.flag ==1)))/
                                      length(which(krig.flag ==1))*100,
```

````
          thr20.non.smoke= (length(which(thresh.20 == 0 & krig.flag ==0)))/
                    length(which(krig.flag ==0))*100,
          thr20.smoke    = (length(which(thresh.20 == 1 & krig.flag ==1)))/
                    length(which(krig.flag ==1))*100,
          thr15.non.smoke= (length(which(thresh.15 == 0 & krig.flag ==0)))/
                    length(which(krig.flag ==0))*100,
          thr15.smoke    = (length(which(thresh.15 == 1 & krig.flag ==1)))/
                    length(which(krig.flag ==1))*100)
gan.compare.gwr2 <- gan.exp.final %>%
 dplyr::summarise(crit.non.smoke = (length(which(final_fire_flag == 0 & gwr.flag ==0)))/
                    length(which(gwr.flag ==0))*100,
          crit.smoke     = (length(which(final_fire_flag == 1 & gwr.flag ==1)))/
                    length(which(gwr.flag ==1))*100,
          site.non.smoke = (length(which(site.spec == 0 & gwr.flag ==0)))/
                    length(which(gwr.flag ==0))*100,
          site.smoke     = (length(which(site.spec == 1 & gwr.flag ==1)))/
                    length(which(gwr.flag ==1))*100,
          thr20.non.smoke= (length(which(thresh.20 == 0 & gwr.flag ==0)))/
                    length(which(gwr.flag ==0))*100,
          thr20.smoke    = (length(which(thresh.20 == 1 & gwr.flag ==1)))/
                    length(which(gwr.flag ==1))*100,
          thr15.non.smoke= (length(which(thresh.15 == 0 & gwr.flag ==0)))/
                    length(which(gwr.flag ==0))*100,
          thr15.smoke    = (length(which(thresh.15 == 1 & gwr.flag ==1)))/
                    length(which(gwr.flag ==1))*100)
gan.compare.wrf2 <- gan.exp.final %>%
 dplyr::summarise(crit.non.smoke = (length(which(final_fire_flag == 0 & wrf.flag ==0)))/
                    length(which(wrf.flag ==0))*100,
          crit.smoke     = (length(which(final_fire_flag == 1 & wrf.flag ==1)))/
                    length(which(wrf.flag ==1))*100,
          site.non.smoke = (length(which(site.spec == 0 & wrf.flag ==0)))/
                    length(which(wrf.flag ==0))*100,
          site.smoke     = (length(which(site.spec == 1 & wrf.flag ==1)))/
                    length(which(wrf.flag ==1))*100,
          thr20.non.smoke= (length(which(thresh.20 == 0 & wrf.flag ==0)))/
                    length(which(wrf.flag ==0))*100,
          thr20.smoke    = (length(which(thresh.20 == 1 & wrf.flag ==1)))/
                    length(which(wrf.flag ==1))*100,
          thr15.non.smoke= (length(which(thresh.15 == 0 & wrf.flag ==0)))/
                    length(which(wrf.flag ==0))*100,
          thr15.smoke    = (length(which(thresh.15 == 1 & wrf.flag ==1)))/
                    length(which(wrf.flag ==1))*100)
```
```{r}
#by method. convert wide to long
````

```
gan.exp.long <- gather(gan.exp.final, method, flag, final_fire_flag, thresh.15:site.spec,
factor_key=F)
gan.exp.long$flag <- factor(gan.exp.long$flag)
gan.exp.long$method <- factor(gan.exp.long$method)
levels(gan.exp.long$method) <- c("Criteria method", "Site specific method", "> 15 ug/m3", ">
20 ug/m3")
levels(gan.exp.long$flag) <- c("Non-smoke day", "Smoke day")
exp.compare <- ggplot(gan.exp.long, aes(x=Conc_ugm3, fill=flag)) +
        facet_wrap(~method) +
        geom_density(alpha=0.5) + xlim(0,100) +
        xlab("PM2.5 (ug/m3)") + ylab("Density") +
        theme_bw() +
          theme(panel.grid.major = element_blank(), panel.grid.minor =
element_blank()) + theme(axis.title.y = element_text(margin = margin(r = 15), size = 12),
          axis.title.x = element_text(margin = margin(t = 15), size = 12),
          legend.title=element_blank(),
          legend.text = element_text(margin = margin(l = 5), size = 10),
          strip.text = element_text(size=12),
          axis.text = element_text(size=10))
exp.compare

#plot each of the gan methods
gan.exp.final$krig.flag <- factor(gan.exp.final$krig.flag)
levels(gan.exp.final$krig.flag) <- c("Non-smoke day", "Smoke day")
gan.krig <- ggplot(gan.exp.final, aes(x=krig.conc, fill=krig.flag))+
    geom_density(alpha=0.5) + xlim(0,100) +
        xlab("Kriged PM2.5 (ug/m3)") + ylab("Density") +
        theme_bw() + theme(panel.grid.major = element_blank(), panel.grid.minor =
element_blank()) + theme(axis.title.y = element_text(margin = margin(r = 15), size = 12),
          axis.title.x = element_text(margin = margin(t = 15), size = 12),
          legend.title=element_blank(),
          legend.text = element_text(margin = margin(l = 5), size = 10),
          strip.text = element_text(size=12),
          axis.text = element_text(size=10))
gan.krig

#########
gan.exp.final$gwr.flag <- factor(gan.exp.final$gwr.flag)
levels(gan.exp.final$gwr.flag) <- c("Non-smoke day", "Smoke day")
gan.gwr <- ggplot(gan.exp.final, aes(x=gwr.conc, fill=gwr.flag))+
    geom_density(alpha=0.5) + xlim(0,100) +
        xlab("GWR PM2.5 (ug/m3)") + ylab("Density") +
        theme_bw() +
          theme(panel.grid.major = element_blank(), panel.grid.minor =
element_blank()) +
          theme(axis.title.y = element_text(margin = margin(r = 15), size = 12),
```

```
                        axis.title.x = element_text(margin = margin(t = 15), size = 12),
                        legend.title=element_blank(),
                        legend.text = element_text(margin = margin(l = 5), size = 10),
                        strip.text = element_text(size=12),
                        axis.text = element_text(size=10))
gan.gwr

#######
gan.exp.final$wrf.flag <- factor(gan.exp.final$wrf.flag)
levels(gan.exp.final$wrf.flag) <- c("Non-smoke day", "Smoke day")
gan.wrf <- ggplot(gan.exp.final, aes(x=wrf.conc, fill=wrf.flag))+
          geom_density(alpha=0.5) + xlim(0,100) +
                  xlab("WRF PM2.5 (ug/m3)") + ylab("Density") +
                  theme_bw() +
                        theme(panel.grid.major = element_blank(), panel.grid.minor =
element_blank()) +
                        theme(axis.title.y = element_text(margin = margin(r = 15), size = 12),
                        axis.title.x = element_text(margin = margin(t = 15), size = 12),
                        legend.title=element_blank(),
                        legend.text = element_text(margin = margin(l = 5), size = 10),
                        strip.text = element_text(size=12),
                        axis.text = element_text(size=10))
gan.wrf
```
```{r bar.plot, echo=F}
#gan.table
smoke.status <- c(0,1,0,1,0,1,0,1)

gan.table.update <- as.data.frame(cbind(gan.table, smoke.status))
setnames(gan.table.update, old=c("V1", "V2", "V3"), new=c("Kriging", "GWR", "wrf"))
gan.table.smoke <-subset(gan.table.update, gan.table.update$smoke.status == 1)
method <- c("crit", "site.spec", "thr20", "thr15")
gan.smoke <- cbind(gan.table.smoke, method)
gan.smoke$method <- factor(gan.smoke$method)
levels(gan.smoke$method) <- c("Criteria", "Site Specific", "20 ug/m3 threshold", "15 ug/m3
threshold")

#convert wide to long
setnames(gan.smoke, old=c("wrf"), new=c("WRF"))
gan.smoke.long <- gather(gan.smoke, gan.method, percent, Kriging:WRF, factor_key=T)
gan.smoke.compare <- ggplot(gan.smoke.long, aes(fill=gan.method, y=percent, x=method)) +
  geom_bar(position="dodge", stat="identity", colour="black", width=0.5) +
  theme_classic() +
  labs(x="", y="Percent in agreement") +
  theme(legend.title = element_blank(),
        axis.title.y = element_text(margin = margin(r = 10), size = 12),
```

```
        axis.text.x = element_text(margin=margin(t=5), size = 10),
        axis.text.y = element_text(margin=margin(r=5), size = 10),
        legend.text = element_text(margin=margin(l=5), size = 10))
gan.smoke.compare

###
gan.table.nonsmoke <- subset(gan.table.update, gan.table.update$smoke.status == 0)
method <- c("crit", "site.spec", "thr20", "thr15")
gan.nonsmoke <- cbind(gan.table.nonsmoke, method)
gan.nonsmoke$method <- factor(gan.nonsmoke$method)
levels(gan.nonsmoke$method) <- c("Criteria", "Site Specific", "20 ug/m3 threshold", "15 ug/m3
threshold")

#convert wide to long
setnames(gan.nonsmoke, old=c("wrf"), new=c("WRF"))
gan.nonsmoke.long <- gather(gan.nonsmoke, gan.method, percent, Kriging:WRF, factor_key=T)
gan.nonsmoke.compare <- ggplot(gan.nonsmoke.long, aes(fill=gan.method, y=percent,
x=method)) +
  geom_bar(position="dodge", stat="identity", colour="black", width=0.5) +
  theme_classic() +
  labs(x="", y="Percent in agreement") +
  theme(legend.title = element_blank(),
        axis.title.y = element_text(margin = margin(r = 10), size = 12),
        axis.text.x = element_text(margin=margin(t=5), size = 10),
        axis.text.y = element_text(margin=margin(r=5), size = 10),
        legend.text = element_text(margin=margin(l=5), size = 10))
gan.nonsmoke.compare
```

Program 3: Aim 2 Data cleaning and preparation

Readin raw geocodes to determine what % are missing by year

````
```{r readin.raw, eval=F}
g.2006 <- fread("Y:/WF_Data/Mortality data/Annie Death Data/geocodes/geocodes.2006.csv")
g.2007 <- fread("Y:/WF_Data/Mortality data/Annie Death Data/geocodes/geocodes.2007.csv")
g.2008 <- fread("Y:/WF_Data/Mortality data/Annie Death Data/geocodes/geocodes.2008.csv")
g.2009 <- fread("Y:/WF_Data/Mortality data/Annie Death Data/geocodes/geocodes.2009.csv")
g.2010 <- fread("Y:/WF_Data/Mortality data/Annie Death Data/geocodes/geocodes.2010.csv")
g.2011 <- fread("Y:/WF_Data/Mortality data/Annie Death
Data/geocodes/DeathGeocode2011.csv")
g.2012 <- fread("Y:/WF_Data/Mortality data/Annie Death
Data/geocodes/DeathGeocode2012.csv")
g.2013 <- fread("Y:/WF_Data/Mortality data/Annie Death
Data/geocodes/DeathGeocode2013.csv")
g.2014 <- fread("Y:/WF_Data/Mortality data/Annie Death
Data/geocodes/DeathGeocode2014.csv")
g.2015 <- fread("Y:/WF_Data/Mortality data/Annie Death
Data/geocodes/DeathGeocode2015.csv")
g.2016 <- fread("Y:/WF_Data/Mortality data/Annie Death
Data/geocodes/DeathGeocodeF2016.csv")
g.2017 <- fread("Y:/WF_Data/Mortality data/Annie Death
Data/geocodes/DeathGeocodeF2017.csv")
#create df unique on certno, with tract_2010 for 2011-2017
geo.2006 <- g.2006 %>% select(dcertnum, tract)
setnames(geo.2006, old=c("dcertnum"), new=c("certno"))
geo.2007 <- g.2007 %>% select(dcertnum, tract)
setnames(geo.2007, old=c("dcertnum"), new=c("certno"))
geo.2008 <- g.2008 %>% select(dcertnum, tract)
setnames(geo.2008, old=c("dcertnum"), new=c("certno"))
geo.2009 <- g.2009 %>% select(certno, tract2000)
setnames(geo.2009, old=c("tract2000"), new=c("tract"))
geo.2010 <- g.2010 %>% select(certno, tract2000)
setnames(geo.2010, old=c("tract2000"), new=c("tract"))
geo.2011 <- g.2011 %>% select(Certno, `2010Tract`)
setnames(geo.2011, old=c("Certno", "2010Tract"), new=c("certno", "tract2010"))
geo.2012 <- g.2012 %>% select(certno, tract2010)
geo.2013 <- g.2013 %>% select(certno, tract2010)
geo.2014 <- g.2014 %>% select(certno, tract2010)
geo.2015 <- g.2015 %>% select(certno, tract2010)
geo.2016 <- g.2016 %>% select(`State File Number`, `Res Geo Census Tract 2010`)
setnames(geo.2016, old=c("State File Number", "Res Geo Census Tract 2010"), new=c("certno",
"tract2010"))
geo.2017 <- g.2017 %>% select(`State File Number`, `Res Geo Census Tract 2010`)
setnames(geo.2017, old=c("State File Number", "Res Geo Census Tract 2010"), new=c("certno",
"tract2010"))
```
````

Readin necessary files: geocoded death files, spatial join file, death file, exposure grid, exposure grid crosswalk

```{r readin, echo=F}
#combine into 1 file
#geo.allyears <- rbind(geo.2006, geo.2007, geo.2008, geo.2009, geo.2010, geo.2011, geo.2012, geo.2013, geo.2014, geo.2015, #geo.2016)
#geo.allyears$X <- NULL
#write.csv(geo.allyears, "geocodes.allyears.csv")
#################################################################
#Read in 2017 geocodes, change names, output for import into gis
geo.2017 <- fread("Y:/WF_Data/Mortality data/Annie Death Data/geocodes/DeathGeocodeF2017.csv")
setnames(geo.2017, old=c("State File Number", "Res Geo Census Block Grp 2000", "Residence Zip Code", "Res Geo Census Tract 2000", "Res Geo Census Block 2000", "Residence Latitude", "Residence Longitude"), new=c("certno", "blockgrp", "geozip", "tract", "block", "lat", "long"))
geo.2017$death.yr <- 2017
geo.2017 <- geo.2017 %>% select(certno, geozip, lat, long, tract, blockgrp, block, death.yr)
fwrite(geo.2017, "Y:/WF_Data/Mortality data/Annie Death Data/geocodes/gis.data/geo.2017.csv")

#ARCGIS JOIN:read in 2017 joined data from gis, then rbind to join_al
geo.2017.join <- fread("Y:/WF_Data/Exposure_grid_data/geocode_17_tract_join.csv")
geo.2017.join <- geo.2017.join %>% select(Join_Count, certno, geozip, lat, long, NAMELSAD10, blockgrp, block, death_yr, FIRST_xcel)
geo.2017.join$tract <- sub(".*Census Tract ", "", geo.2017.join$NAMELSAD10)
setnames(geo.2017.join, old=c("FIRST_xcel"), new=c("xcell_ycell"))

#drop if join_count ==0
geo.2017.join <- subset(geo.2017.join, geo.2017.join$Join_Count==1)
geo.2017.join <- geo.2017.join %>% select(certno, death_yr, blockgrp, geozip, tract, block, lat, long, xcell_ycell, Join_Count)
setnames(geo.2017.join, old=c("death_yr"), new=c("deathyr"))
####################################################################
#readin spatial join file
library(data.table)

#setwd("Y:/WF_Data/Mortality data/Annie Death Data/geocodes/gis output")
#join_all_old <- fread("Y:/WF_Data/Mortality data/Annie Death Data/geocodes/gis output/final_spatial_join.csv")

#drop and rename variables
#join_all_subset <- join_all %>%
#  select("certno", "deathyr", "blockgrp", "geozip", "tract", "block", "lat", "long", "FIRST_xcel", "Join_Count")
#setnames(join_all_subset, old=c("FIRST_xcel"), new=c("xcell_ycell"))
```

join_all <- fread("Y:/WF_Data/Exposure_grid_data/geocode_06_16_tract_join_update.csv")
join_all_subset <- join_all %>%
  dplyr::select("certno", "deathyr", "blockgrp", "geozip", "NAMELSAD10", "block", "lat",
"long", "FIRST_xcel", "Join_Count")
join_all_subset$tract <- sub(".*Census Tract ", "", join_all_subset$NAMELSAD10)

setnames(join_all_subset, old=c("FIRST_xcel"), new=c("xcell_ycell"))
join_all_subset <- subset(join_all_subset, join_all_subset$Join_Count==1)
join_all_subset <- join_all_subset %>% select(certno, deathyr, blockgrp, geozip, tract, block, lat,
long, xcell_ycell, Join_Count)

#rbind w/2017 data; order columns
join_all_complete <- rbind(join_all_subset, geo.2017.join)
fwrite(join_all_complete, "Y:/WF_Data/Mortality data/Annie Death
Data/combined/join_all_complete.csv")
```
```{r read.in.all, echo=F}
join_all_complete <- fread("Y:/WF_Data/Mortality data/Annie Death
Data/combined/join_all_complete.csv")
combine.death <- fread("Y:/WF_Data/Mortality data/Annie Death
Data/death.files/combined.death.files.csv")

#read in exposure data
exposure.grid <-
fread("Y:/WF_Data/Exposure_grid_data/Population_by_grid_cells_w_fire_flags_FINAL.csv")
setnames(exposure.grid, old=c('site_assigned'), new=c('AQS_ID'))

##need site variable. need to create xwalk of aqs_id and site name to merge onto exp grid, then
merge this onto death data
exp.grid.xwalk <- fread("Y:/WF_Data/Exposure_grid_data/AQS.ID_Monitor_xwalk.csv")

#read in county code file
county.code <- fread("Y:/WF_Data/SES/median_income/county.code.csv")

#read in tract/county/median household income file
mhi.trct <- fread("Y:/WF_Data/SES/median_income/tract.medinc.csv")
#setnames(mhi.trct, old=c("year"), new=c("deathyr"))

```
Merge on spatial join file on combined death data to get all death data characteristics. Then
merge on exposure day (by date and grid cell), keeping all. Then, for each death, assign referent
days within each calendar month by day of the week. For each death, we'll have 1 row for their
event, and a row for each referent day
```{r combine}
setnames(county.code, old=c("code"), new=c("cnty_occ"))
county.code <- select(county.code, cnty_occ, county)

```
#merge spatial data with death file
merge.all <- merge(join_all_complete, combine.death, by = 'certno', all=T)
#merge on county code to assign county names to codes in death file
merge.all.county <- merge(merge.all, county.code, by="cnty_occ", all=T)
merge.all.county$tract <- as.numeric(merge.all.county$tract)
merge.all.county$year <- substr(merge.all.county$date, 1, 4)
merge.all.county$year <- as.numeric(merge.all.county$year)

#Merge on median household income data by census tract, county, and year
merge.all.mhi <- merge(merge.all.county, mhi.trct, by=c("tract", "county", "year"), all=T)

#keep only deaths in months between 6 and 9, according to our inclusion criteria (June-
September)
merge.all.summer <- subset(merge.all.mhi, merge.all.mhi$dth_mo > 5 & merge.all.mhi$dth_mo
<= 9)
#rename df
merge.all.complete <- merge.all.summer

#keep only non-traumatic deaths - look at underly, which is the underlying cause of death, which
we'll use to define cause of death
merge.all.complete$underly.letter <- substr(merge.all.complete$underly, 1, 1)
merge.all.complete$underly.num <- as.numeric(substr(merge.all.complete$underly, 2, 4))
#drop traumatic (V, W, X, Y)
merge.all.complete <- subset(merge.all.complete, underly.letter %in% c("A", "B", "C", "D", "E",
"F", "G", "H", "I", "J", "K", "L", "M", "N", "O", "P", "Q", "R"))
merge.all.complete$mlt1.letter <- substr(merge.all.complete$mltcse1, 1, 1)
merge.all.complete$mlt1.num <- as.numeric(substr(merge.all.complete$mltcse1, 2, 4))

#label cause of death: cardiovascular and respirator
merge.all.complete$cardio <- ifelse((merge.all.complete$underly.letter == "I" &
merge.all.complete$underly.num < 530
                              & merge.all.complete$underly.num > 40) |
                               (merge.all.complete$mlt1.letter == "I" &
                                merge.all.complete$mlt1.num < 530 &
                                 merge.all.complete$mlt1.num > 40) , 1, 0)
merge.all.complete$resp <- ifelse((merge.all.complete$underly.letter == "J") |
(merge.all.complete$mlt1.letter == "J"), 1, 0)
#asthma: J45 and J46.
merge.all.complete$asthma <- with(merge.all.complete, ifelse((underly.letter == "J" &
underly.num %in% c(450, 459)) | (mlt1.letter == "J" &
                                          mlt1.num %in% c(450, 459)), 1, 0))
#copd: J41-44
merge.all.complete$copd <- with(merge.all.complete,
                               ifelse((underly.letter == "J" &
                                     underly.num %in% c(432, 439, 440, 441, 448, 449)) |
```

```
                          (mlt1.letter == "J" & mlt1.num %in% c(432, 439, 440, 441, 448, 449)),
1, 0))
#pneumonia
merge.all.complete$pneumonia <- with(merge.all.complete,
                         ifelse((underly.letter == "J" &
                                underly.num %in% c(120, 129, 150, 151, 152, 154, 156, 157, 158,
                                                   159, 180, 181, 182, 188, 189)) |
                              (mlt1.letter == "J" &
                                mlt1.num %in% c(120, 129, 150, 151, 152, 154, 156, 157, 158,
                                                159, 180, 181, 182, 188, 189)), 1, 0))
#Ischemic heart disease
merge.all.complete$heart.disease <- with(merge.all.complete,
                            ifelse((underly.letter == "I" &
                                underly.num %in% c(200, 201, 208, 209, 212, 214, 219,
                                                     229, 248, 249)) |
                                (mlt1.letter == "I" &
                                mlt1.num %in% c(200, 201, 208, 209, 212, 214, 219,
                                                  229, 248, 249)), 1, 0))
#Cerebrovascular causes of death (stroke)
merge.all.complete$cerebrovascular <- with(merge.all.complete,
                             ifelse((underly.letter == "I" &
                                underly.num %in% c(600, 606, 607, 608, 609, 610, 611,
                                                    612, 613, 614, 615, 618, 619, 620, 621,
                                                      629, 630, 631, 632, 633, 634, 635,
                                                      638, 639, 670, 671, 672, 673, 674,
                                                      675, 677, 678, 679)) |
                                 (mlt1.letter == "I" &
                                        mlt1.num %in% c(600, 606, 607, 608, 609, 610, 611,
                                                   612, 613, 614, 615, 618, 619, 620, 621,
                                                   629, 630, 631, 632, 633, 634, 635,
                                                   638, 639, 670, 671, 672, 673, 674,
                                                   675, 677, 678, 679)), 1, 0))

#merge updated death file on exposure crosswalk to get monitoring site names
merge.all.nt.xwalk <- merge(merge.all.complete, exp.grid.xwalk, by=c('xcell_ycell', 'date'), all.x
= T)
#Create dummy variable for urban sites. Includes sites in Seattle area, Tacoma area, and Spokane
merge.all.nt.xwalk$urban <- ifelse(merge.all.nt.xwalk$Site %in% c("Seattle 10th Weller",
"Seattle Beacon Hill",
                                            "Seattle Duwamish", "Seattle South Park", "Bellevue",
                                            "Kent", "Lynnwood", "Lake Forest Park", "Tacoma
Alex",
                                            "Tacoma L", "Tacoma 36", "Puyallup", "Spokane
Broadway",
                                            "Spokane Colbert", "Spokane College", "Spokane
Monroe",
```

~~~~
                                                        "Spokane Augusta", "Airway Heights"), 1, 0)
#create age group categories according to DOH categories
merge.all.nt.xwalk$age.grp <- with(merge.all.nt.xwalk, ifelse(ageunum < 5, 1, ifelse(ageunum
>= 5 & ageunum <15, 2,
                                               ifelse(ageunum >= 15 & ageunum < 45, 3,
                                               ifelse(ageunum >= 45 & ageunum < 65, 4,
                                               ifelse(ageunum >= 65 & ageunum < 85, 5,
                                               ifelse(ageunum >= 85, 6, 0)))))))
#create race category according to DOH categories
merge.all.nt.xwalk$race.cat <- with(merge.all.nt.xwalk, ifelse(race == "1", "White",
                                                ifelse(race == "2", "Black",
                                                ifelse(race == "3", "Native American",
                                                ifelse(race == "C", "Hispanic",
                                                ifelse(race %in% c("A", "B", "F", "H"), "NHOPI",
                                                ifelse(race %in% c("4", "5", "7", "D", "E","G"), "Asian",
                                                ifelse(race == 6, "Other", "Not reported"))))))))
#Create numerical race category
merge.all.nt.xwalk$race.cat.num <- with(merge.all.nt.xwalk, ifelse(race.cat == "White", 1,
                                                    ifelse(race.cat == "Black", 2,
                                                    ifelse(race.cat == "Native American", 3,
                                                    ifelse(race.cat == "Hispanic", 4,
                                                    ifelse(race.cat == "NHOPI", 5,
                                                    ifelse(race.cat == "Asian", 6, 0)))))))
#create education categories
merge.all.nt.xwalk$educ_level <- with(merge.all.nt.xwalk, ifelse(educ >= 1 & educ <= 3, "High
school or less",
                                                  ifelse(educ == 4 | educ == 5, "Some college",
                                                  ifelse(educ == 6, "College",
                                                  ifelse(educ == 7 | educ == 8, "Graduate or advanced degree",
                                                          "Not reported")))))
fwrite(merge.all.nt.xwalk, "Y:/WF_Data/Mortality data/Annie Death
Data/combined/death.nontrauma.csv")
```
Calculate the humidex spline using the splines package.
```{r spline}
#add a 20ug/m3 cutpoint
exposure.grid$thresh.20 <- ifelse(exposure.grid$Conc_ugm3 >= 20, 1, 0)

#create concentration per 10 unit change; divide by 10
exposure.grid$conc.10 <- exposure.grid$Conc_ugm3/10

#fitting a linear spline for humidex. natural cubic spline with 3 degrees of freedom
exposure.grid$humidexSp <- ns(exposure.grid$Humidex, df=3)
```
Case-crossover datafile preparation.
Merge our files, then create referent days for each case.
~~~~

```{r casecross}
#merge death data on exposure data, by date and grid cell, keeping all
final.merged.death.file <- merge(merge.all.nt.xwalk, exposure.grid, by=c("date", "xcell_ycell"),
all=T)

#prepare case crossover dataset
#need to create several rows for each case: one row for the event, and a row for each referrent
day
final.merged.death.file$case <- ifelse(is.na(final.merged.death.file$certno),0,1)

#Create new df where case == 1
df.cases <- subset(final.merged.death.file, final.merged.death.file$case ==1)

#keep only certno and date
df.cases.subset <- df.cases %>%
  select(certno, date, xcell_ycell, final_fire_flag, cause.death, cardio, resp, asthma, copd,
pneumonia, heart.disease, cerebrovascular, med_inc)

df.cases.subset$month <- month(df.cases.subset$date)
df.case.subset <- subset(df.cases.subset, df.cases.subset$month >5 & df.cases.subset$month <10)
df.cases.subset$month <- NULL

#for each case, identify the set of referent days (for now, do the 3 weeks before and 3 weeks
after)
df.cases.subset$ref.day1 <- as.factor(ymd(df.cases.subset$date) - 28)
df.cases.subset$ref.day2 <- as.factor(ymd(df.cases.subset$date) - 21)
df.cases.subset$ref.day3 <- as.factor(ymd(df.cases.subset$date) - 14)
df.cases.subset$ref.day4 <- as.factor(ymd(df.cases.subset$date) - 7)
df.cases.subset$ref.day5 <- as.factor(ymd(df.cases.subset$date) + 7)
df.cases.subset$ref.day6 <- as.factor(ymd(df.cases.subset$date) + 14)
df.cases.subset$ref.day7 <- as.factor(ymd(df.cases.subset$date) + 21)
df.cases.subset$ref.day8 <- as.factor(ymd(df.cases.subset$date) + 28)
df.cases.subset$index.month <- month(df.cases.subset$date)

setnames(df.cases.subset, old=c("date", "ref.day1", "ref.day2", "ref.day3", "ref.day4", "ref.day5",
"ref.day6", "ref.day7", "ref.day8"), new=c("lag0.index", "lag0.ref.day1", "lag0.ref.day2",
"lag0.ref.day3", "lag0.ref.day4", "lag0.ref.day5", "lag0.ref.day6", "lag0.ref.day7",
"lag0.ref.day8"))

fwrite(df.cases.subset, "Y:/WF_Data/Annie_CaseCross/df.cases.subset.csv")

#convert wide to long
df.cases.long <- melt(df.cases.subset, id.vars=c("certno", "xcell_ycell", "index.month"),
measure.vars=c("lag0.index", "lag0.ref.day1", "lag0.ref.day2", "lag0.ref.day3", "lag0.ref.day4",
"lag0.ref.day5", "lag0.ref.day6", "lag0.ref.day7", "lag0.ref.day8"), variable.name = "day",
value.name = "date")
```

````
#get rid of referent days in May and October
df.cases.long$month <- month(df.cases.long$date)
df.referents <- subset(df.cases.long, df.cases.long$month >5 & df.cases.long$month <10)

#assign case days and control days
df.referents$case <- ifelse(df.referents$day == "lag0.index", 1, 0)

#keep only referent days within the same month as the case day
df.ref.select <- subset(df.referents, df.referents$month == df.referents$index.month)

#create lag day 1 date variable in order to merge on exposure grid by lag 1 date and xcell ycell to get associated concs and humidex
#create 2 datasets: one w/dropped 6/1 and one without

df.ref.select$lag1.date <- as.factor(ymd(df.ref.select$date)-1)
```
Create final case crossover dataset
```{r create.final}
#CREATE FINAL DF
####################
#merge the df.referents on the exposure grid. merge on date and xcell_ycell
merge.experiment <- merge(df.ref.select, exposure.grid, by=c("date", "xcell_ycell"))

#merge this on the death file to get all the individual variables; merge on certno
final.case.cross <- merge(merge.experiment, merge.all.nt.xwalk, by=c("certno"))

setnames(final.case.cross, old="cnty_occ", new="code")

#final.case <- merge(final.case.cross, mhi.county, by=c("year", "code", "tract"), all=T)

setnames(final.case.cross, old=c("final_fire_flag", "xcell_ycell.x", "date.x"),
        new=c("smoke_day", "xcell_ycell", "date"))

final.case.cross$humidexSp <- ns(final.case.cross$Humidex, df=3)
setnames(final.case.cross, old="year.x", new="year")

final.subset <- final.case.cross %>% dplyr::select(case, certno, ageunum, smoke_day,
Conc_ugm3, race.cat.num, cause.death, urban, date, Humidex, resp, cardio, asthma, copd,
age.grp, year, humidexSp, thresh.20, pneumonia, heart.disease, cerebrovascular, educ, med_inc)

fwrite(final.subset, "Y:/WF_Data/Annie_CaseCross/case.cross.analysis.file.csv")

##########################################################################
#CREATE FINAL DF W/LAG1 EXPOSURE INCLUDED
#merge the df.referents on the exposure grid. merge on lag1.date and xcell_ycell
````

~~~~
merge.experiment.lag1 <- subset(merge.experiment, merge.experiment$date != "2006-06-01")

merge.experiment.lag1$humidexSp <- ns(merge.experiment.lag1$Humidex, df=3)

exp.lag1 <- select(merge.experiment.lag1, certno, xcell_ycell, index.month, day, date, month,
case, lag1.date, Conc_ugm3, Humidex, humidexSp, final_fire_flag)

setnames(exp.lag1, old=c("Conc_ugm3", "Humidex", "humidexSp", "final_fire_flag", "date"),
         new=c("Conc_ugm3.lag0", "Humidex.lag0", "humidexSp.lag0",
"smoke_day.lag0","lag0.date"))

pm.grid.lag1 <- exposure.grid %>% dplyr::select(xcell_ycell, date, year, Conc_ugm3, Humidex,
final_fire_flag)
setnames(pm.grid.lag1, old=c("date"), new=c("lag1.date"))

merge.exp.lag1 <- merge(exp.lag1, pm.grid.lag1, by=c("lag1.date", "xcell_ycell"))

#merge this on the death file to get all the individual variables; merge on certno
final.case.cross.lag1 <- merge(merge.exp.lag1, merge.all.nt.xwalk, by=c("certno"))
final.case.cross.lag1$humidexSp <- ns(final.case.cross.lag1$Humidex, df=3)

setnames(final.case.cross.lag1, old=c("xcell_ycell.x", "Conc_ugm3", "Humidex", "humidexSp",
"final_fire_flag"),
                new=c("xcell_ycell", "Conc_ugm3.lag1", "Humidex.lag1",
"humidexSp.lag1", "smoke_day.lag1"))

final.subset.lag1 <- final.case.cross.lag1 %>% dplyr::select(case, certno, ageunum,
smoke_day.lag0, smoke_day.lag1, Conc_ugm3.lag0, Conc_ugm3.lag1, race.cat.num,
cause.death, urban, date, humidexSp.lag0, humidexSp.lag1, Humidex.lag0, Humidex.lag1, resp,
cardio, asthma, copd, lag1.date, lag0.date, age.grp, pneumonia, heart.disease, cerebrovascular,
educ)

fwrite(final.subset.lag1, "Y:/WF_Data/Annie_CaseCross/case.cross.lag1.analysis.file.csv")
```
~~~~

Create exposure matrix for lag analysis, for lag days 0-4.

~~~~
```{r lag1, eval=F}
#assign referrent days for lag days 1 to 4
#lag day 1 + referrent days
df.cases.subset <- fread("Y:/WF_Data/Annie_CaseCross/df.cases.subset.csv", header=T, sep=',')
df.cases.subset$index.month <- month(df.cases.subset$lag0.index)
df.cases.subset$index.day <- day(df.cases.subset$lag0.index)
df.cases.subset$ref1.day <- day(df.cases.subset$lag0.ref.day1)
df.cases.subset$ref2.day <- day(df.cases.subset$lag0.ref.day2)
df.cases.subset$ref3.day <- day(df.cases.subset$lag0.ref.day3)
df.cases.subset$ref4.day <- day(df.cases.subset$lag0.ref.day4)
df.cases.subset$ref5.day <- day(df.cases.subset$lag0.ref.day5)
~~~~

```
df.cases.subset$ref6.day <- day(df.cases.subset$lag0.ref.day6)
df.cases.subset$ref7.day <- day(df.cases.subset$lag0.ref.day7)
df.cases.subset$ref8.day <- day(df.cases.subset$lag0.ref.day8)

#drop deaths on first 4 days of June; and drop referent days on first 4 days of June
df.cases.subset$drop <- with(df.cases.subset, ifelse(index.month == 6 & (index.day %in%
c(1,2,3,4) |
                                        ref1.day %in% c(1,2,3,4) |
                                        ref2.day %in% c(1,2,3,4) |
                                        ref3.day %in% c(1,2,3,4) |
                                        ref4.day %in% c(1,2,3,4) |
                                        ref5.day %in% c(1,2,3,4) |
                                        ref6.day %in% c(1,2,3,4) |
                                        ref7.day %in% c(1,2,3,4) |
                                        ref8.day %in% c(1,2,3,4)), 1, 0))
#drop cases and referents on first 4 days of June
df.cases.sub <- subset(df.cases.subset, drop==0)

#rename df
df.cases.lag <- df.cases.sub

#----------------------Lag 1 referent days-------------------------------------
#create lag1 index day and referent days
df.cases.lag$lag1.index <- as.factor(ymd(df.cases.lag$lag0.index)-1)
df.cases.lag$lag1.month <- month(df.cases.lag$lag1.index)
#create referent days for lag day 1
df.cases.lag$lag1.ref.day1 <- as.factor(ymd(df.cases.lag$lag1.index) - 28)
df.cases.lag$lag1.ref.day2 <- as.factor(ymd(df.cases.lag$lag1.index) - 21)
df.cases.lag$lag1.ref.day3 <- as.factor(ymd(df.cases.lag$lag1.index) - 14)
df.cases.lag$lag1.ref.day4 <- as.factor(ymd(df.cases.lag$lag1.index) - 7)
df.cases.lag$lag1.ref.day5 <- as.factor(ymd(df.cases.lag$lag1.index) + 7)
df.cases.lag$lag1.ref.day6 <- as.factor(ymd(df.cases.lag$lag1.index) + 14)
df.cases.lag$lag1.ref.day7 <- as.factor(ymd(df.cases.lag$lag1.index) + 21)
df.cases.lag$lag1.ref.day8 <- as.factor(ymd(df.cases.lag$lag1.index) + 28)
df.cases.lag$lag1.index <- as.factor(df.cases.lag$lag1.index)

#----------------------Lag 2 referent days-------------------------------------
#lag day 2 + referrent days
df.cases.lag$lag2.index <- as.factor(ymd(df.cases.lag$lag0.index) - 2)
df.cases.lag$lag2.ref.day1 <- as.factor(ymd(df.cases.lag$lag2.index) - 28)
df.cases.lag$lag2.ref.day2 <- as.factor(ymd(df.cases.lag$lag2.index) - 21)
df.cases.lag$lag2.ref.day3 <- as.factor(ymd(df.cases.lag$lag2.index) - 14)
df.cases.lag$lag2.ref.day4 <- as.factor(ymd(df.cases.lag$lag2.index) - 7)
df.cases.lag$lag2.ref.day5 <- as.factor(ymd(df.cases.lag$lag2.index) + 7)
df.cases.lag$lag2.ref.day6 <- as.factor(ymd(df.cases.lag$lag2.index) + 14)
df.cases.lag$lag2.ref.day7 <- as.factor(ymd(df.cases.lag$lag2.index) + 21)
```

~~~
df.cases.lag$lag2.ref.day8 <- as.factor(ymd(df.cases.lag$lag2.index) + 28)
df.cases.lag$lag2.index <- as.factor(df.cases.lag$lag2.index)

#----------------------Lag 3 referent days-------------------------------------
#lag day 3 + referrent days
df.cases.lag$lag3.index <- as.factor(ymd(df.cases.lag$lag0.index) - 3)
df.cases.lag$lag3.ref.day1 <- as.factor(ymd(df.cases.lag$lag3.index) - 28)
df.cases.lag$lag3.ref.day2 <- as.factor(ymd(df.cases.lag$lag3.index) - 21)
df.cases.lag$lag3.ref.day3 <- as.factor(ymd(df.cases.lag$lag3.index) - 14)
df.cases.lag$lag3.ref.day4 <- as.factor(ymd(df.cases.lag$lag3.index) - 7)
df.cases.lag$lag3.ref.day5 <- as.factor(ymd(df.cases.lag$lag3.index) + 7)
df.cases.lag$lag3.ref.day6 <- as.factor(ymd(df.cases.lag$lag3.index) + 14)
df.cases.lag$lag3.ref.day7 <- as.factor(ymd(df.cases.lag$lag3.index) + 21)
df.cases.lag$lag3.ref.day8 <- as.factor(ymd(df.cases.lag$lag3.index) + 28)
df.cases.lag$lag3.index <- as.factor(df.cases.lag$lag3.index)

#----------------------Lag 4 referent days-------------------------------------
#lag day 4 + referrent days
df.cases.lag$lag4.index <- as.factor(ymd(df.cases.lag$lag0.index) - 4)
df.cases.lag$lag4.ref.day1 <- as.factor(ymd(df.cases.lag$lag4.index) - 28)
df.cases.lag$lag4.ref.day2 <- as.factor(ymd(df.cases.lag$lag4.index) - 21)
df.cases.lag$lag4.ref.day3 <- as.factor(ymd(df.cases.lag$lag4.index) - 14)
df.cases.lag$lag4.ref.day4 <- as.factor(ymd(df.cases.lag$lag4.index) - 7)
df.cases.lag$lag4.ref.day5 <- as.factor(ymd(df.cases.lag$lag4.index) + 7)
df.cases.lag$lag4.ref.day6 <- as.factor(ymd(df.cases.lag$lag4.index) + 14)
df.cases.lag$lag4.ref.day7 <- as.factor(ymd(df.cases.lag$lag4.index) + 21)
df.cases.lag$lag4.ref.day8 <- as.factor(ymd(df.cases.lag$lag4.index) + 28)
df.cases.lag$lag4.index <- as.factor(df.cases.lag$lag4.index
```
Create exposure matrix with case-crossover layout
```{r lag2, eval=F}
#Load nt death file
merge.all.nt.xwalk <- fread("Y:/WF_Data/Mortality data/Annie Death
Data/combined/death.nontrauma.csv", header=T, sep=',')

#----------------------Lag day 1 reshape-------------------------------------
#select certno
merge.xwalk <- merge.all.nt.xwalk %>% select("certno")

#reshape wide to long to create case-crossover dataset
df.cases.long.lag1 <- melt(df.cases.lag, id.vars=c("certno", "index.month", "lag0.index",
"xcell_ycell"), measure.vars=c("lag1.index", "lag1.ref.day1", "lag1.ref.day2", "lag1.ref.day3",
"lag1.ref.day4", "lag1.ref.day5", "lag1.ref.day6", "lag1.ref.day7", "lag1.ref.day8"), variable.name
= "day", value.name = "date")

#get rid of referent days in May and October
~~~

```
df.cases.long.lag1$month <- month(df.cases.long.lag1$date)
df.referents.lag1 <- subset(df.cases.long.lag1, df.cases.long.lag1$index.month >5 &
df.cases.long.lag1$index.month <10)

#assign case days and control days
df.referents.lag1$case <- ifelse(df.referents.lag1$day == "lag1.index", 1, 0)
df.referents.lag1$index.day <- day(df.referents.lag1$lag0.index)

#keep only referent days within the same month as the case day unless index day on first four
days of month; keep lag1.index
df.referents.lag1$keep <- ifelse(df.referents.lag1$month == df.referents.lag1$index.month |
                              (df.referents.lag1$index.month %in% c(7,8,9) &
df.referents.lag1$index.day %in% c(1,2,3,4)), 1, 0)
df.ref.select.lag1 <- subset(df.referents.lag1, df.referents.lag1$keep == 1)

merge.lag1 <- merge(df.ref.select.lag1, exposure.grid, by=c("date", "xcell_ycell"))
#merge.lag1.final <- merge(merge.lag1, merge.all.nt.xwalk, by=c("certno"))

lag1 <- merge.lag1 %>% select(certno, date, case, Conc_ugm3, humidexSp, lag0.index)

#----------------------Lag day 2 reshape-------------------------------------
df.cases.long.lag2 <- melt(df.cases.lag, id.vars=c("certno", "xcell_ycell", "index.month",
"lag0.index"), measure.vars=c("lag2.index", "lag2.ref.day1", "lag2.ref.day2", "lag2.ref.day3",
"lag2.ref.day4", "lag2.ref.day5", "lag2.ref.day6", "lag2.ref.day7", "lag2.ref.day8"), variable.name
= "lag.day2", value.name = "date")

#get rid of referent days in May and October
df.cases.long.lag2$month <- month(df.cases.long.lag2$date)
df.referents.lag2 <- subset(df.cases.long.lag2, df.cases.long.lag2$index.month >5 &
df.cases.long.lag2$index.month <10)

#assign case days and control days
df.referents.lag2$case <- ifelse(df.referents.lag2$lag.day2 == "lag2.index", 1, 0)
df.referents.lag2$index.day <- day(df.referents.lag2$lag0.index)

#keep only referent days within the same month as the case day unless index day on first four
days of month; keep lag1.index
df.referents.lag2$keep2 <- ifelse(df.referents.lag2$month == df.referents.lag2$index.month |
                              (df.referents.lag2$index.month %in% c(7,8,9) &
df.referents.lag2$index.day %in% c(1,2,3,4)), 1, 0)
df.ref.select.lag2 <- subset(df.referents.lag2, df.referents.lag2$keep2 == 1)

merge.lag2 <- merge(df.ref.select.lag2, exposure.grid, by=c("date", "xcell_ycell"))
#merge.lag2.final <- merge(merge.lag2, merge.all.nt.xwalk, by=c("certno"))
lag2 <- merge.lag2 %>% select(certno, date, case, Conc_ugm3, humidexSp, lag0.index)
```

```
#---------------------Lag day 3 reshape------------------------------------
df.cases.long.lag3 <- melt(df.cases.lag, id.vars=c("certno", "xcell_ycell", "index.month",
"lag0.index"), measure.vars=c("lag3.index", "lag3.ref.day1", "lag3.ref.day2", "lag3.ref.day3",
"lag3.ref.day4", "lag3.ref.day5", "lag3.ref.day6", "lag3.ref.day7", "lag3.ref.day8"), variable.name
= "lag.day3", value.name = "date")

#get rid of referent days in May and October
df.cases.long.lag3$month <- month(df.cases.long.lag3$date)
df.referents.lag3 <- subset(df.cases.long.lag3, df.cases.long.lag3$index.month >5 &
df.cases.long.lag3$index.month <10)

#assign case days and control days
df.referents.lag3$case <- ifelse(df.referents.lag3$lag.day3 == "lag3.index", 1, 0)
df.referents.lag3$index.day <- day(df.referents.lag3$lag0.index)

#keep only referent days within the same month as the case day unless index day on first four
days of month; keep lag1.index
df.referents.lag3$keep3 <- ifelse(df.referents.lag3$month == df.referents.lag3$index.month |
                              (df.referents.lag3$index.month %in% c(7,8,9) &
df.referents.lag3$index.day %in% c(1,2,3,4)), 1, 0)
df.ref.select.lag3 <- subset(df.referents.lag3, df.referents.lag3$keep3 == 1)

merge.lag3 <- merge(df.ref.select.lag3, exposure.grid, by=c("date", "xcell_ycell"))
#merge.lag3.final <- merge(merge.lag3, merge.all.nt.xwalk, by=c("certno"))
lag3 <- merge.lag3 %>% select(certno, date, case, Conc_ugm3, humidexSp, lag0.index)

#---------------------Lag day 4 reshape-------------------------------------
df.cases.long.lag4 <- melt(df.cases.lag, id.vars=c("certno", "xcell_ycell", "index.month",
"lag0.index"), measure.vars=c("lag4.index", "lag4.ref.day1", "lag4.ref.day2", "lag4.ref.day3",
"lag4.ref.day4", "lag4.ref.day5", "lag4.ref.day6", "lag4.ref.day7", "lag4.ref.day8"), variable.name
= "lag.day4", value.name = "date")

#get rid of referent days in May and October
df.cases.long.lag4$month <- month(df.cases.long.lag4$date)
df.referents.lag4 <- subset(df.cases.long.lag4, df.cases.long.lag4$index.month >5 &
df.cases.long.lag4$index.month <10)

#assign case days and control days
df.referents.lag4$case <- ifelse(df.referents.lag4$lag.day4 == "lag4.index", 1, 0)
df.referents.lag4$index.day <- day(df.referents.lag4$lag0.index)

#keep only referent days within the same month as the case day unless index day on first four
days of month
df.referents.lag4$keep4 <- ifelse(df.referents.lag4$month == df.referents.lag4$index.month |
                              (df.referents.lag4$index.month %in% c(7,8,9) &
df.referents.lag4$index.day %in% c(1,2,3,4)), 1, 0)
```

```
df.ref.select.lag4 <- subset(df.referents.lag4, df.referents.lag4$keep4 == 1)

merge.lag4 <- merge(df.ref.select.lag4, exposure.grid, by=c("date", "xcell_ycell"))
#merge.lag4.final <- merge(merge.lag4, merge.all.nt.xwalk, by=c("certno"))
lag4 <- merge.lag4 %>% select(certno, date, case, Conc_ugm3, humidexSp, lag0.index)

df.cases.long <- melt(df.cases.sub, id.vars=c("certno", "xcell_ycell", "index.month"),
measure.vars=c("lag0.index", "lag0.ref.day1", "lag0.ref.day2", "lag0.ref.day3", "lag0.ref.day4",
"lag0.ref.day5", "lag0.ref.day6", "lag0.ref.day7", "lag0.ref.day8"), variable.name = "day",
value.name = "date")

#get rid of referent days in May and October
df.cases.long$month <- month(df.cases.long$date)
df.cases.long$index.day <- day(df.cases.long$date)
df.referents <- subset(df.cases.long, df.cases.long$month >5 & df.cases.long$month <10)

#assign case days and control days
df.referents$case <- ifelse(df.referents$day == "lag0.index", 1, 0)

#keep only referent days within the same month as the case day unless index day on first four
days of month; keep lag1.index
df.referents$keep0 <- ifelse(df.referents$month == df.referents$index.month |
                          (df.referents$index.month %in% c(7,8,9) & df.referents$index.day %in%
c(1,2,3,4)), 1, 0)
df.ref.select <- subset(df.referents, df.referents$keep0 == 1)

#merge the df.referents on the exposure grid. merge on date and xcell_ycell
merge.experiment <- merge(df.ref.select, exposure.grid, by=c("date", "xcell_ycell"))

merge.lag0 <- merge.experiment
lag0 <- merge.lag0 %>% select(certno, date, case, Conc_ugm3, humidexSp, final_fire_flag)
setnames(lag0, old=c('final_fire_flag'), new=c('smoke_day'))

#merge all lag days (0-4)
#sort by certno and case, add case number so unique (1-4, with 1 being case), and merge on
certno and case)
lag0 <- lag0[order(lag0$certno, -lag0$case),]
lag0.sort <- ddply(lag0, .(certno), mutate, case.num = seq_along(case))
setnames(lag0.sort, old=c("Conc_ugm3", "date", "humidexSp"), new=c("conc.lag0", "date.lag0",
"humidex.lag0"))
lag0.sort$case <- NULL

lag1 <- lag1[order(lag1$certno, -lag1$case),]
setnames(lag1, old=c('date', 'case', 'Conc_ugm3', 'humidexSp'), new=c('date.lag1', 'case.lag1',
'conc.lag1', 'humidex.lag1'))
lag1.sort <- ddply(lag1, .(certno), mutate, case.num = seq_along(case.lag1))
```

```
lag1.sort$case.lag1 <- NULL
lag1.sort$lag0.index <- NULL

lag2 <- lag2[order(lag2$certno, -lag2$case),]
lag2.sort <- ddply(lag2, .(certno), mutate, case.num = seq_along(case))
setnames(lag2.sort, old=c("date", "Conc_ugm3", "humidexSp"), new=c("date.lag2", "conc.lag2",
"humidex.lag2"))
lag2.sort$case <- NULL
lag2.sort$lag0.index <- NULL

lag3 <- lag3[order(lag3$certno, -lag3$case),]
lag3.sort <- ddply(lag3, .(certno), mutate, case.num = seq_along(case))
setnames(lag3.sort, old=c("date", "Conc_ugm3", "humidexSp"), new=c("date.lag3", "conc.lag3",
"humidex.lag3"))
lag3.sort$case <- NULL
lag3.sort$lag0.index <- NULL

lag4 <- lag4[order(lag4$certno, -lag4$case),]
lag4.sort <- ddply(lag4, .(certno), mutate, case.num = seq_along(case))
setnames(lag4.sort, old=c("date", "Conc_ugm3", "humidexSp"), new=c("date.lag4", "conc.lag4",
"humidex.lag4"))
lag4.sort$case <- NULL
lag4.sort$lag0.index <- NULL

#merge/join
merge1 <- merge(lag0.sort, lag1.sort, by=c("certno", "case.num"), all=T)
merge2 <- merge(merge1, lag2.sort, by=c("certno", "case.num"))
merge3 <- merge(merge2, lag3.sort, by=c("certno", "case.num"))
merge4 <- merge(merge3, lag4.sort, by=c("certno", "case.num"))

#drop vars so we just have certno, case.num, and concentrations; renumber case.num
merge4$case <- ifelse(merge4$case.num > 1, 0, 1)

exp.matrix <- merge4 %>%
  select(certno, case, smoke_day, conc.lag0, humidex.lag0, conc.lag1, humidex.lag1, conc.lag2,
humidex.lag2, conc.lag3, humidex.lag3, conc.lag4, humidex.lag4)

#merge on df.cases.subset by certno to get cause of death vars; select to keep only vars we care
about
df.cause.death <- df.cases.subset %>% dplyr::select(certno, cause.death, cardio, resp, asthma,
copd)
merge.exp.matrix <- merge(exp.matrix, df.cause.death, by=c("certno"))
fwrite(merge.exp.matrix, "Y:/WF_Data/Annie_CaseCross/lag.exp.matrix.csv")
```

Program 4: Final manuscript table creation

```
---
```{r readin, echo=F}
#create function to output summary info
output.summary <- function(model, tablenum ="Table1", tablename = "Table") {
  output <- round(exp(cbind(OR = coef(model), confint(model))), 3)
  p <- round(summary(model)$coefficients[,5], 3)
  sig <- ifelse(p<=0.001, "***", ifelse((p<=0.01),"**", ifelse((p<=0.05),"*",
ifelse((p<=0.1),".",""))))
  model.results <- data.frame(output, p, sig)
  setnames(model.results, old=c('X2.5..', 'X97.5..'), new=c('Lower 95% CI', 'Upper 95% CI'))

  table.out <- kable(model.results, "latex", booktabs = T) %>%
    kable_styling(latex_options = c("striped", "hold_position")) %>%
    footnote(general = tablename, general_title = tablenum, title_format=c("bold"),
footnote_as_chunk = T)
  sum.out <- summary(model)
  return(table.out)
  #return(sum.out)
}
```
```{r table3.models, echo=F}
###Model 1.1
model1.1 <- clogit(case~smoke_day + humidexSp + strata(certno), final.case.cross)

output1.1 <- round(exp(cbind(OR = coef(model1.1), confint(model1.1))), 3)
p1.1 <- round(summary(model1.1)$coefficients[,5], 3)
sig1.1 <- ifelse(p1.1<=0.001, "***", ifelse((p1.1<=0.01),"**", ifelse((p1.1<=0.05),"*",
                                            ifelse((p1.1<=0.1),".",""))))
model.results1.1 <- data.frame(output1.1, p1.1, sig1.1)
setnames(model.results1.1, old=c('X2.5..', 'X97.5..'), new=c('Lower 95% CI', 'Upper 95% CI'))

### Model 1.2
##Ages 0-4##
model1.2A <- clogit(case~smoke_day + humidexSp + strata(certno), data=final.case.cross,
                                          subset = age.grp == 1)
##Ages 5-14##
model1.2B <- clogit(case~smoke_day + humidexSp + strata(certno), data=final.case.cross,
                                          subset = age.grp == 2)
##Ages 15-44##
model1.2C <- clogit(case~smoke_day + humidexSp + strata(certno), data=final.case.cross,
                                          subset = age.grp == 3)
##Ages 45-64##
```
```

```
model1.2D <- clogit(case~smoke_day + humidexSp + strata(certno), data=final.case.cross,
                                        subset = age.grp == 4)
##Ages 65-84##
model1.2E <- clogit(case~smoke_day + humidexSp + strata(certno), data=final.case.cross,
                                        subset = age.grp == 5)
##Ages 85+##
model1.2F <- clogit(case~smoke_day + humidexSp + strata(certno), data=final.case.cross,
                                        subset = age.grp == 6)

OR.bin.summ <- function(model) {
  coef  <- as.data.frame(coef(model))
  conf  <- as.data.frame(confint(model))
  OR    <- round(exp(coef[1, 1]), 2)
  LCL   <- round(exp(conf[1, 1]), 2)
  UCL   <- round(exp(conf[1, 2]), 2)
  p     <- round(summary(model)$coef[1,5], 2)
  sig   <- ifelse(p<0.001, "***", ifelse((p<0.01),"**", ifelse((p<0.05),"*", ifelse((p<0.1),".",""))))
  table <- cbind(OR, LCL, UCL, p, sig)
  return(table)
}

age1 <- OR.bin.summ(model1.2A)
age2 <- OR.bin.summ(model1.2B)
age3 <- OR.bin.summ(model1.2C)
age4 <- OR.bin.summ(model1.2D)
age5 <- OR.bin.summ(model1.2E)
age6 <- OR.bin.summ(model1.2F)

Age <- c("0-4", "5-14", "15-44", "45-64", "65-84", "85+")

table1.2 <- rbind(age1, age2, age3, age4, age5, age6)
table.age <- cbind(Age, table1.2)

###Model 1.3 - cause of death
##Cause: cardio##
model1.3A <- clogit(case~smoke_day + humidexSp + strata(certno), data = final.case.cross,
                                        subset = cardio == 1)
##Cause: respiratory##
model1.3B <- clogit(case~smoke_day + humidexSp + strata(certno), data = final.case.cross,
                                        subset = resp == 1)
model1.3C <- clogit(case~smoke_day + humidexSp + strata(certno), data = final.case.cross,
                                        subset = asthma == 1)
model1.3D <- clogit(case~smoke_day + humidexSp + strata(certno), data = final.case.cross,
```

```
                                          subset = copd == 1)
model1.3E <- clogit(case~smoke_day + humidexSp + strata(certno), data = final.case.cross,
                                          subset = pneumonia == 1)
model1.3F <- clogit(case~smoke_day + humidexSp + strata(certno), data = final.case.cross,
                                          subset = heart.disease == 1)
model1.3G <- clogit(case~smoke_day + humidexSp + strata(certno), data = final.case.cross,
                                          subset = cerebrovascular == 1)

cause1 <- OR.bin.summ(model1.3A)
cause2 <- OR.bin.summ(model1.3B)
cause3 <- OR.bin.summ(model1.3C)
cause4 <- OR.bin.summ(model1.3D)
cause5 <- OR.bin.summ(model1.3E)
cause6 <- OR.bin.summ(model1.3F)
cause7 <- OR.bin.summ(model1.3G)
all <- OR.bin.summ(model1.1)

Cause <- c("Cardiovascular", "All respiratory", "Asthma", "COPD", "Pneumonia", "Ischemic
heart disease",
        "Cerebrovascular", "All non-traumatic")

table1.3 <- rbind(cause1, cause2, cause3, cause4, cause5, cause6, cause7, all)
table.cause <- cbind(Cause, table1.3)

###
##Stratification by race
#Binary exposure with threshold as the criteria smoke day, stratifying by race category, and
adjusting for humidex.
##Race: White##
model1.4A <- clogit(case~smoke_day + humidexSp + strata(certno), data = final.case.cross,
                                          subset = race.cat.num == 1)
##Race: Black
model1.4B <- clogit(case~smoke_day + humidexSp + strata(certno), data = final.case.cross,
                                          subset = race.cat.num == 2)
#Race: Native American
model1.4C <- clogit(case~smoke_day + humidexSp + strata(certno), data = final.case.cross,
                                          subset = race.cat.num == 3)
#Race: Hispanic
model1.4D <- clogit(case~smoke_day + humidexSp + strata(certno), data = final.case.cross,
                                          subset = race.cat.num == 4)
#Race: NHOPI
model1.4E <- clogit(case~smoke_day + humidexSp + strata(certno), data = final.case.cross,
                                          subset = race.cat.num == 5)
```

```
#Race: Asian
model1.4F <- clogit(case~smoke_day + humidexSp + strata(certno), data = final.case.cross,
                                        subset = race.cat.num == 6)
race1 <- OR.bin.summ(model1.4A)
race2 <- OR.bin.summ(model1.4B)
race3 <- OR.bin.summ(model1.4C)
race4 <- OR.bin.summ(model1.4D)
race5 <- OR.bin.summ(model1.4E)
race6 <- OR.bin.summ(model1.4F)

Race <- c("White", "Black", "Native American", "Hispanic", "NHOPI", "Asian")
table1.4 <- rbind(race1, race2, race3, race4, race5, race6)
table.race <- cbind(Race, table1.4)

###SES
model.SES1 <- clogit(case~smoke_day + humidexSp + strata(certno), data = final.case.cross,
                                        subset = house_income == 1)
model.SES2 <- clogit(case~smoke_day + humidexSp + strata(certno), data = final.case.cross,
                                        subset = house_income == 2)
model.SES3 <- clogit(case~smoke_day + humidexSp + strata(certno), data = final.case.cross,
                                        subset = house_income == 3)
model.SES4 <- clogit(case~smoke_day + humidexSp + strata(certno), data = final.case.cross,
                                        subset = house_income == 4)
model.SES5 <- clogit(case~smoke_day + humidexSp + strata(certno), data = final.case.cross,
                                        subset = house_income == 5)

SES1 <- OR.bin.summ(model.SES1)
SES2 <- OR.bin.summ(model.SES2)
SES3 <- OR.bin.summ(model.SES3)
SES4 <- OR.bin.summ(model.SES4)
SES5 <- OR.bin.summ(model.SES5)

SES <- c("<35k", "35-50k", "50-75k", "75-100k", ">100k")

tableSES <- rbind(SES1, SES2, SES3, SES4, SES5)
table.SES <- cbind(SES, tableSES)

##Stratification by rural/urban
#Urban vs. rural stratification, exposured modeled as binary, with threshold as the criteria smoke
day, adjusting for humidex

#Urban
model1.5A <- clogit(case~smoke_day + humidexSp + strata(certno), data = final.case.cross,
```

```
                                             subset = urban == 1)
#Rural
model1.5B <- clogit(case~smoke_day + humidexSp + strata(certno), data = final.case.cross,
                                             subset = urban == 0)
urban <- OR.bin.summ(model1.5A)
rural <- OR.bin.summ(model1.5B)

Location <- c("Urban", "Rural")
table1.5 <- rbind(urban, rural)
table.location <- cbind(Location, table1.5)
```

Figure 1. Distributed lag model

```
```{r fig1.out, echo=F}
exp.matrix <- fread("/home/gradstudent/doubleda/My
Documents/Thesis/analysis/lag.exp.matrix.update.csv")
exp.matrix$conc10.lag0 <- exp.matrix$conc.lag0/10
exp.matrix$conc10.lag1 <- exp.matrix$conc.lag1/10
exp.matrix$conc10.lag2 <- exp.matrix$conc.lag2/10
exp.matrix$conc10.lag3 <- exp.matrix$conc.lag3/10
exp.matrix$conc10.lag4 <- exp.matrix$conc.lag4/10

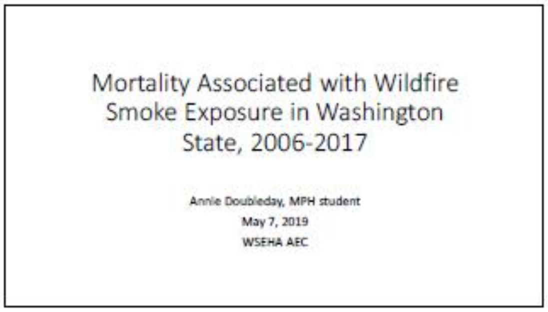
model3.0 <- clogit(case~smoke_day + conc10.lag0 + conc10.lag1 + conc10.lag2 + conc10.lag3
+ conc10.lag4 + humidex.lag0 + humidex.lag1 + humidex.lag2 + humidex.lag3 + humidex.lag4
+ strata(certno), exp.matrix)

#output.summary(model3.0, tablenum = "Table 13.", tablename = "Lag regression output")
confint <- exp(confint(model3.0))
conc.ci <- confint[2:6,]
coef <- as.data.table(exp(coef(model3.0)))
conc.coef <- coef[2:6,]
lag <- c("Lag 0", "Lag 1", "Lag 2", "Lag 3", "Lag 4")
coefs.plot <- cbind(conc.coef, conc.ci, lag)
setnames(coefs.plot, old=c("V1", "2.5 %", "97.5 %"), new=c("OR", "LCL", "UCL"))

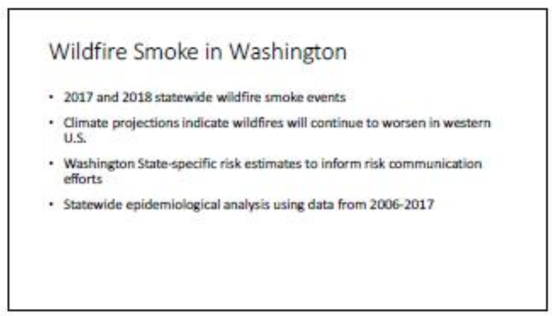
p <-  ggplot(data=coefs.plot, aes(x = lag, y = OR, ymin = LCL, ymax = UCL))+
    geom_pointrange()+ ylim(0.95, 1.05) +
    geom_hline(aes(fill=lag),yintercept =1, linetype=1)+
    labs(title = "Figure 1. Lag model Odds Ratios, days 0-4",
        x = "", y = "Odds Ratio (95% Confidence Interval)") +
    geom_errorbar(aes(ymin=LCL, ymax=UCL), width=0.1, cex=0.8) +
    theme_classic()
p
```
```

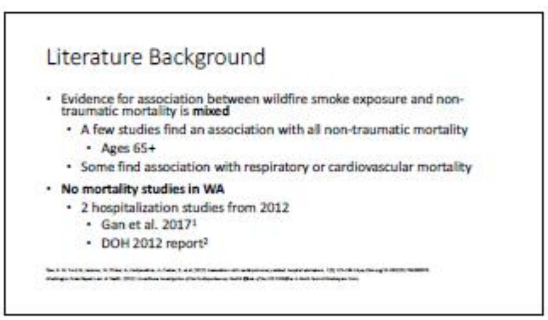
## Appendix E. Research Translation Materials

## WSEHA Slides






# Methods

## Developing wildfire smoke day definition

- Goal: develop a wildfire smoke day definition for the Washington State
- Challenge
  - Minimize contribution of high anthropogenic $PM_{2.5}$ days in urban areas while still capturing wildfire smoke influenced days across the state
- Methods considered
  - $PM_{2.5}$ thresholds: 15, 20 $\mu g/m^3$
  - Site-specific threshold
  - **Criteria-based method (final method)**

## Methods

- Exposure metric
  - 24-hour average $PM_{2.5}$ and humidex data from air quality monitoring network checked against modeled $PM_{2.5}$ from the AIRPACT model
- Health outcome
  - Non-traumatic mortality in Washington State
  - June-September, 2006-2017
- Time-stratified case-crossover design
  - Cases compared to themselves to control for confounding
  - Compare wildfire smoke exposure on day of death to exposure on referent days
  - Estimate odds ratio day of exposure, and on lag days 1-4

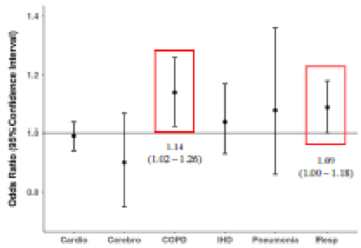
## Hypotheses

- No change in odds of non-traumatic mortality for wildfire smoke days compared to non-wildfire smoke days, controlling for humidex
  - Small increase in odds for cardiovascular and respiratory mortality
- Increase in odds of non-traumatic mortality for ages 65+, on wildfire smoke days compared to non-wildfire smoke days, controlling for humidex
- Increase in odds of non-traumatic mortality for all ages at a lag of 1 day

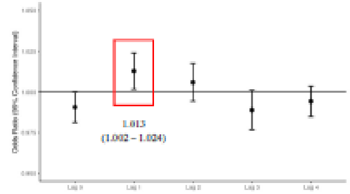
# Results

## Results: Primary Analysis

| Category | Adjusted OR (95% CI) | N (%) |
|---|---|---|
| All non-traumatic | 1.01 (0.99, 1.04) | 171,804 (100) |
| Age group (years) | | |
| 65-84 | 1.02 (0.98, 1.06) | 75,110 (43.4) |
| 85+ | 1.00 (0.96, 1.05) | 57,618 (33.3) |
| Cause of death | | |
| Cardiovascular | 0.99 (0.94, 1.04) | 44,565 (25.9) |
| Respiratory | 1.09 (1.00, 1.18) | 16,286 (9.5) |




## Results: Lag analysis




## Results: Secondary Analyses

| Category | Adjusted OR (95% CI) | N (%) |
|---|---|---|
| All non-traumatic | 1.01 (0.99, 1.04) | 171,804 (100) |
| Respiratory causes, by age group | | |
| 0-4 | 1.52 (0.58, 3.97) | 119 (0.7) |
| 5-14 | - | 28 (0.2) |
| 15-44 | 0.91 (0.45, 1.84) | 217 (1.3) |
| 45-64 | 1.35 (1.09, 1.67)* | 2,152 (13.2) |
| 65-84 | 1.08 (0.96, 1.21) | 8,489 (52.1) |
| 85+ | 1.00 (0.86, 1.16) | 5,281 (32.4) |
| Age group, for COPD causes of death | | |
| 45-64 | 1.33 (1.00, 1.78) | 1,281 (13.4) |
| 65-84 | 1.14 (0.99, 1.31) | 5,654 (59.1) |
| 85+ | 1.04 (0.85, 1.28) | 2,584 (27.0) |

# Discussion

## Discussion

- Overall effect estimates from primary analysis are similar to what other studies have found for non-traumatic mortality
- No other studies have examined COPD mortality
  - Evidence in literature for association between wildfire smoke exposure and an increase in COPD morbidity
- No other studies find statistically significant increase in any cause of mortality for ages 45-64
  - Possible explanation: older worker effect
- Lagged effect seen in many studies

## Limitations

- Challenge of separating anthropogenic $PM_{2.5}$ and wildfire smoke $PM_{2.5}$
- $PM_{2.5}$ area monitors
  - Assuming area exposure equals personal exposure
- Inclusion of 2018 data would increase power and tighten confidence intervals of effect estimates

## Conclusion

- Evidence for some association between wildfire smoke exposure and non-traumatic mortality among specific populations and specific causes of death
- Research needed to determine a gold standard method for wildfire smoke exposure
- Research needed to further explore subgroups, to examine possible associations between other health outcomes of interest, and to explore long term exposure
- Effect estimates will help inform state agency risk communication efforts for wildfire seasons to come

## Thank you!

Questions?

Annie Doubleday, MPH Student

doubleda@uw.edu

W ENVIRONMENTAL & OCCUPATIONAL HEALTH SCIENCES
UNIVERSITY of WASHINGTON | SCHOOL OF PUBLIC HEALTH

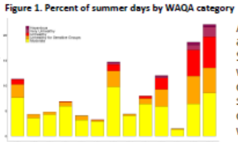
## NEHA Poster

# Mortality Associated with Wildfire Smoke Exposure in Washington State

Annie Doubleday[1], Jill Schulte[2], Ranil Dhammapala[2], Matt Kadlec[2], Julie Fox[3], Lianne Sheppard[1], Tania Busch Isaksen[1]
[1]Department of Environmental and Occupational Health Sciences, University of Washington
[2]Air Quality Program, Washington State Department of Ecology
[3]Environmental Public Health Sciences, Washington State Department of Health

## BACKGROUND

Figure 1. Percent of summer days by WAQA category

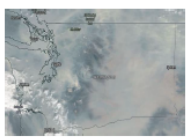


Air quality designated as Unhealthy for Sensitive Groups, or worse, has increased during recent summers in WA state due to increased wildfires in the state and region[1] (Fig. 1).

Figure 2. $PM_{2.5}$ summer 2017

Daily $PM_{2.5}$ concentrations vary spatially, with areas closer to the fires typically experiencing higher $PM_{2.5}$ than areas further from the source (Fig. 2).

## METHODS



NASA MODIS Worldview

The association between wildfire smoke exposure and non-traumatic mortality was evaluated using a time-stratified case-crossover study design.

Wildfire smoke affected days were defined as days above 20.4 μg/m³, and through an additional set of criteria for days between 9 and 20.4 μg/m³.

Each case was assigned a $PM_{2.5}$ concentration from the nearest air quality monitor on the day of death and on 3 control days. Exposure on the death date and control dates were compared.

Conditional logistic regression was used to evaluate the associations. Results are reported as a change in the odds of mortality on wildfire smoke days vs. non-wildfire smoke days.

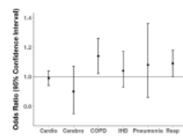
## RESULTS

Table 1. ORs (95% CIs) of non-traumatic mortality

| Category | | Adjusted OR (95% CI) |
|---|---|---|
| All non-traumatic | | 1.01 (0.99, 1.04) |
| **Age group (years)** | | |
| | 0-4 | 0.97 (0.77, 1.22) |
| | 5-14 | 0.95 (0.64, 1.41) |
| | 15-44 | 0.99 (0.85, 1.15) |
| | 45-64 | 1.00 (0.95, 1.06) |
| | 65-84 | 1.02 (0.98, 1.06) |
| | 85+ | 1.00 (0.96, 1.05) |
| **Median Household income by census tract** | | |
| | Non-urban | 1.00 (0.96, 1.03) |
| | Urban | 1.02 (0.99,1.06) |
| | <$35,000 | 0.99 (0.92, 1.07) |
| | $35,000 - $50,000 | 1.05 (0.99, 1.10) |
| | $50,000 - $75,000 | 1.00 (0.96, 1.04) |
| | $75,000 - $100,000 | 1.03 (0.96, 1.10) |
| | ≥$100,000 | 0.98 (0.87, 1,11) |
| **Race** | | |
| | White | 1.01 (0.99, 1.04) |
| | Black | 1.04 (0.89, 1.21) |
| | Native American | 0.95 (0.77, 1.17) |
| | Hispanic | 0.81 (0.67, 0.99)* |
| | Native Hawaiian/ Other Pacific Islander | 1.19 (0.93, 1.52) |
| | Asian | 0.98 (0.84, 1.14) |
| **Sex** | | |
| | Female | 1.00 (0.96, 1.03) |
| | Male | 1.03 (0.99, 1.06) |
| **Underlying cause of death** | | |
| | Cardiovascular | 0.99 (0.94, 1.04) |
| | Ischemic heart disease | 1.04 (0.93, 1.17) |
| | Respiratory | 1.09 (1.00, 1.18) |
| | COPD | 1.14 (1.02, 1.26)* |
| | Pneumonia | 1.08 (0.86, 1.36) |
| | Cerebrovascular | 0.90 (0.75, 1.07) |

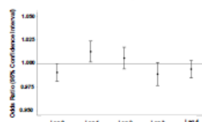
Results in Table 1 indicate a small non-significant increase in all same-day non-traumatic mortality on wildfire smoke days compared to non-wildfire smoke days, controlling for humidex (measure of heat). Positive effect estimates among specific groups: median household income $35-50k; males; respiratory and COPD mortality are also presented.

Figure 3. ORs (95% CIs) by Cause of Death




Evidence for an effect among respiratory and COPD mortality, with little evidence for an effect among other causes of death (Fig. 3).

Figure 4. ORs (95% CIs) by Lag Day from an unconstrained lag model




Evidence for a lagged effect on non-traumatic mortality from wildfire smoke exposure on the two days preceding the outcome (Fig. 4).

## DISCUSSION

The overall effect estimate for non-traumatic mortality is similar to what other studies have found.[2]

Wikipedia: Carlton Complex Fire 2014

We find evidence for an effect among COPD mortality. No other studies have examined COPD mortality.[2]

We find evidence for an effect among adults 65 and over, as well as adults 45-64 for respiratory mortality. No other studies report an effect in any cause of mortality for ages 45-64.[2]

Limitations

Separating background anthropogenic and wildfire smoke $PM_{2.5}$ is a challenge when defining a wildfire smoke event.

Use of $PM_{2.5}$ area monitors assumes area exposure is equal to personal exposure, and does not take into account indoor air exposure.

Low sample size among some sensitive populations yields inadequate power to detect an effect.

## CONCLUSION

There is evidence suggesting an association between wildfire smoke exposure and mortality, among adults 45 and over and for respiratory mortality.

Effect estimates will help inform state agency risk communication efforts for upcoming wildfire seasons.

Further Research and Next Steps

- Inclusion of 2018 data will increase power and tighten confidence intervals of effect estimates.
- Exposure assessment with measured and modeled $PM_{2.5}$ would improve accuracy of personal $PM_{2.5}$ exposure.
- Need for gold standard exposure assessment method.
- Examine possible associations for other health outcomes of interest in Washington, and further examine subgroups.

*References:*
*1. Abatzoglou, J. T., & Williams, A. P. (2016). Impact of anthropogenic climate change on wildfire across western US forests. https://doi.org/10.1073/pnas.1607171113*
*2. Reid et al. 2016. Critical review of health impacts of wildfire smoke exposure. Environ Health Persp.*

*Acknowledgements:*
*Washington State Department of Health for use of mortality data*
*Washington State Department of Ecology for use of exposure data*

ENVIRONMENTAL & OCCUPATIONAL HEALTH SCIENCES
UNIVERSITY of WASHINGTON | SCHOOL OF PUBLIC HEALTH