Challenges of Using Publicly-Available Hospital Data to Quantify Health Effects from Wildfire Plumes in the East San Francisco Bay Area Communities of California, USA

Natasha Atkins[*a] *& Ronald L. Baskett*[b]

[a]*College of Agricultural & Environmental Sciences, University of California, Davis, CA*
[b]*Tri-Valley Air Quality Climate Alliance, Livermore, CA*

https://doi.org/10.33697/ajur.2024.111

*Student: naatkins@formerstudents.ucdavis.edu**
Mentor: ronbaskett@tvaqca.org

ABSTRACT
In the summer and fall of 2018 and 2020, major wildfires in Northern California (USA) impacted the San Francisco Bay Area. The remote 2018 and 2020 wildfires produced the highest $PM_{2.5}$ concentrations experienced in the Tri-Valley of the East Bay Area during those two years. The Tri-Valley is composed of the San Ramon, Amador, and Livermore Valleys, surrounded by local terrain that creates a small airshed. In 1967, the California Air Resources Board created 15 Air Basins defined by regional geography, topographic and meteorological conditions. Airshed is sometimes synonymous with an urban-scale component of an Air Basin. We use airshed as a Tri-Valley component of the Bay Area Air Basin. This airshed spans across two counties (Contra Costa and Alameda) and encompasses four cities: San Ramon, Dublin, Pleasanton, and Livermore. PurpleAir (PA) sensors provided good geographic coverage of variation in $PM_{2.5}$ in the Tri-Valley airshed. Several studies have established significant health effects from wildfire plumes by associating daily hospital visits with $PM_{2.5}$ air quality data at local and regional scales. We hypothesize that during the wildfire smoke periods of 2018 and 2020 in the Tri-Valley area, there was an increase in hospitalizations and ED visits for respiratory (asthma and COPD-related) health effects, as compared to the same time periods during years with less fire activity. The primary goal of this study was to confirm health effects from wildfire plumes on a community scale using 5 years of publicly-available health data. However, with only monthly hospitalization data available, directly linking respiratory hospital and emergency department (ED) visits with $PM_{2.5}$ concentrations was unsuccessful. Also, because COVID-19 masked all other causes of hospital visits in 2020, that year was ultimately eliminated from this study. However, visits during November 2018 being much higher than any other November in 2016, 2017, and 2019 implied a potential cause and effect. Daily hospitalization and air quality data are required to quantify any relationship by regression analysis. These findings help inform future studies on the health effects of air quality at community scales.

KEYWORDS
$PM_{2.5}$ Air Quality; Air Pollutant Exposure; Air Quality Monitoring; Wildfire Smoke; Respiratory Health; PurpleAir Sensors

INTRODUCTION

Exposure to wildfire smoke very likely increases acute respiratory health effects.[1-3] The negative effects associated with wildfire smoke exposure have become an increasingly pressing issue as California experiences more wildfires. In 2020 alone 3.6 million acres were burned in California and residents were exposed to smoke for many more days than normal.[3] Our study focuses on the San Francisco East Bay Area's Tri-Valley, an airshed within Contra Costa and Alameda counties. The geographic nature (surrounding foothills that can trap pollutants in the valleys during stagnant wind conditions)[1] of Tri-Valley creates the potential for wildfire smoke and associated air pollutants, particularly fine particulate matter ($PM_{2.5}$, particles that are 2.5 microns or smaller in diameter) to settle in the valleys where the cities of San Ramon, Dublin, Pleasanton, and Livermore are located.

In recent years (2018 and 2020), there have been instances where the Tri-Valley has been affected by an influx of wildfire smoke from Northern California wildfires. In 2018, wildfire smoke from the Camp Fire was likely present on 13 of the 15 days when the 24-hour $PM_{2.5}$ federal standard of 35 $\mu g/m^3$ was exceeded in the Tri-Valley that year. The Camp Fire was the "most deadly and destructive"[4] wildfire in California, occurring in the town of Paradise on November 8, 2018, and was fully contained by November 25, 2018.[5] The fire burned 153,336 acres, 19,356 structures, and resulted in 85 deaths.[5] Specific meteorological conditions made the Camp Fire severe. With speeds over 20 mph, hot, dry downslope winds known as Diablo winds caused the fire to not only spread rapidly in Paradise, but they also carried the smoke plume into the Tri-Valley airshed.[6]

In 2020, the Santa Clara Unit (SCU) Lightning Fire Complex burned at the end of summer into fall. The complex of 20 fires first started on August 16 and wasn't fully contained until October 1.[7] The fires burned 396,624 acres,[7] with the two largest fires merging and burning south of Livermore and Pleasanton.[8] In 2020, all 17 $PM_{2.5}$ exceedance days (days where the national standard was exceeded, see explanation in the following paragraph) in the Tri-Valley were probably due to smoke from the Santa Clara Unit (SCU) Lightning Fire Complex entering the area.

The Bay Area Air Quality Management District (Air District) recommends staying indoors during these unhealthy smoke days. Previous studies have found that while indoor $PM_{2.5}$ concentrations are typically lower than outdoor concentrations, indoor concentrations still follow outdoor increases.[9, 10] The Environmental protection agency (EPA) regulates air quality nationwide through their Air Quality System (AQS).[11] The EPA also establishes National Ambient Air Quality Standards (NAAQS), with the annual primary standard for $PM_{2.5}$ being 9 $\mu g\ m^{-3}$.[12]

Low-cost sensors can be used to measure air pollutants at higher spatial and temporal scales than regulatory agency monitoring networks with less operating skill and significantly less cost. While the key advantage of low-cost sensors is their low expense, the major disadvantage is their accuracy. The South Coast AQMD Air Quality Sensor Performance Evaluation Center (AQ-SPEC) reviewed sensor characteristics and performance. In their 2016laboratory and field comparisons, both Purple Air PA-I and PA-II $PM_{2.5}$ sensors were found to correlate with Federal Reference Method (FRM) or Federal Equivalent Method (FEM) sensors with linear correlation coefficients greater than 0.9 for 5-minute data as well as 1-hour averages.[13] Both AQ-SPEC and Barkjohn, et al. (2021) found that the PurpleAir Sensors had good precision but read high by approximately 40%.[13, 14] Wallace, et al. (2022) found that PA-I and PA-II measurements can agree well with regulatory monitors when an optimum calibration factor of 3.4 is applied.[15]

PA, a network of low-cost indoor and outdoor air quality sensors, has been used to observe spatial trends in $PM_{2.5}$ concentrations.[16] The advantages of using the PA sensor network for this study are that there are many sensors that record publicly available data. The number of sensors allows for high spatial coverage of the Tri-Valley area. However, some of the disadvantages of PA are that sensors are set up by the public and inconsistencies in sensor placement can affect data quality. Additionally, sensors can have unreliable/irregular recording, with no or incorrect recordings for periods of time. Widespread use of PA sensors is relatively new, and many sensors were not up and running until after 2018. For these reasons the PA data quality is not the same as quality assured air district monitors (see: http://www.aqmd.gov/aq-spec/home) Comparatively, district sensors (part of the EPA AQS) follow EPA requirements and the data is quality assured and controlled.[11] EPA AQS monitoring stations are likely more accurate than PA sensors for $PM_{2.5}$ data; however, there are more PA sensors than district sensors with 1,171 out of 5,824 census tracts in California having one to two fully operational PA sensors.[17] Compared to the approximately 250 sensors in the California Ambient Air Monitoring Network.[18]

Previous studies have addressed whether or not wildfire smoke and associated $PM_{2.5}$ have significant effects on (respiratory) health outcomes.[2, 19, 20] Recent epidemiological studies have shown that $PM_{2.5}$ from wildfire smoke can exacerbate a range of health problems including respiratory and cardiovascular issues.[2, 19, 21-23] Several studies link wildfire-specific $PM_{2.5}$ exposures to increases in respiratory hospitalizations. Other studies further aim to determine which groups are most affected,[19, 24] while others aim to determine "whether $PM_{2.5}$ from wildfires is more or less harmful [to human health] than PM2.5 from other sources" [25] For instance, Aguilera *et al.* (2021) found that $PM_{2.5}$ from wildfire smoke events in Southern California was up to ten times worse for human health than non-wildfire $PM_{2.5}$ in terms of respiratory hospital admissions.[25]

Another topic of interest in previous studies is the spatial aspect of respiratory health effects;[24, 16] one these studies, Southerland *et al.* (2022), specifically aims to determine if exposure to $PM_{2.5}$ at smaller (city) scales shows a different trend in $PM_{2.5}$ attributable mortality than larger spatial scale studies.[16] This information can give us a better understanding of spatial hotspots of negative health outcomes related to $PM_{2.5}$ exposure, which would allow for targeted mitigation strategies. Typically, the variation in health impacts at finer spatial scales is masked because impacts are generally reported at the country, state, or county level.[16] So, Southerland *et al.* (2022) studied how air pollution-related health risks vary at the neighborhood scale within cities, specifically Bay Area cities, to assist with public health decision making.[16] The researchers determined that pollution-attributable risks can vary considerably within individual cities.[16] Furthermore, neighborhoods with the highest air pollution-attributable health risks were not necessarily those with the highest pollutant concentrations.[16] However, there were limitations to the study in that baseline disease rates are hard to obtain at urban and intra-urban scales.[16] So, for asthma incidence the researchers were only able to use the statewide incidence rate, despite the fact that there is data evidence for spatial variation in asthma incidence in the Bay Area.[16]

A focus of recent studies, relevant to our study, is indoor vs outdoor $PM_{2.5}$ monitoring and exposure. As O'Dell *et al.* (2022) point out, most air quality research is focused on outdoor pollutant concentrations, while people tend to spend more of their time indoors.[10] Using publicly-available data from low-cost sensors, O'Dell *et al.* (2022) found that during periods of heavy smoke, indoor $PM_{2.5}$ concentrations could exceed the 35 μg m^{-3} 24-hour federal outdoor standard.[10] Also, they found a "total daily-mean indoor $PM_{2.5}$ concentrations increase by 2.1 μg m^{-3} with every 10-μg m^{-3} increase in daily-mean outdoor $PM_{2.5}$."[10] For our study, we utilize PA sensors to look at both indoor and outdoor concentrations across the Tri-Valley. We compare one nearby indoor and outdoor PA sensor to each of the two district sensors during the 2020 wildfire period (Table 1), in order to gain a better understanding of indoor $PM_{2.5}$ concentrations during a wildfire smoke episode, given the amount of time people generally spend indoors.[10]

Some researchers conducting epidemiological studies associated with wildfire smoke have access to non-publicly accessible hospitalization data. For example, Alexeeff *et al.* (2023) used a cohort of 3.7 million adults in the Kaiser Permanente Northern California health care system.[26] From this large dataset they were able to determine that long-term $PM_{2.5}$ exposure was associated with cardiovascular health issues.[26] Both Delfino *et al.* and Aguilera *et al.* obtained their hospitalization data from the California Office of Statewide Health Planning and Development, however, this data was not accessible for this research.[19, 25] One important thing to note about most studies (aside from Alexeeff *et al.*)[26] is that they look at acute health effects resulting from exposure to only one smoke event, rather than considering cumulative health effects from multiple exposure events.[27] This means its possible health impacts are more extensive than reported in these single-event studies.[27] Even with some studies having great accessibility to data, a review by Black *et al.* (2017) notes that there must be a more standardized consensus on both air quality monitoring and reporting during wildfire events in order to allow for more thorough epidemiological, between-study comparisons.[28]

In this paper we aim to gain a better understanding of one, how PA sensor data compare to quality-assured air district monitoring data during normal air quality days as well as days when a wildfire smoke plume is present in the Tri-Valley; and two, the potential respiratory health effects of the 2018 and 2020 wildfire smoke plumes present in the Tri-Valley. Our basic hypothesis is that increases in wildfire smoke $PM_{2.5}$ are correlated with increases in hospitalizations and ED visits for asthma and COPD-related health effects during the 2018 Camp Fire and 2020 SCU Lightning Fire Complex wildfire periods.
We explore this hypothesis with several questions:

1) Are $PM_{2.5}$ concentrations within the Tri-Valley significantly different when a wildfire smoke plume is present versus when one is not?
2) Are PA sensors as accurate as district sensors during typical air quality days and during wildfire smoke days?
3) Is there spatial variation in $PM_{2.5}$ concentrations within the Tri-Valley during a wildfire smoke plume?
4) Can a regression analysis of monthly hospitalization and ED data and monthly $PM_{2.5}$ concentrations be used to evaluate wildfire effects on respiratory health, and if so, is there a significant increase in respiratory hospitalizations and ED visits during wildfire smoke plumes?

METHODS AND PROCEDURES

The geographic area of focus for this paper is the Tri-Valley Airshed, encompassing the cities of Livermore, Pleasanton, and San Ramon, and Dublin as shown in **Figure 1**. The PA sensor network in the Tri-Valley Airshed was used in combination with Air District sensors, which are a part of the EPA AQS,[11] to evaluate changes in $PM_{2.5}$ concentrations across wildfire years and years with less wildfire activity from 2016 to 2020, and potential associations with respiratory hospitalizations and emergency department (ED) visits. The goal was to determine how exposure to $PM_{2.5}$ from the 2018 Camp Fire and the 2020 Lightning Fire Complex impacted the health of Tri-Valley residents.

$PM_{2.5}$ and hospitalization data

The period of study from 2016 to 2020 encompasses the Camp Fire in 2018 and the SCU Lightning Complex Fires in 2020. $PM_{2.5}$ data was collected from two sources. First, quality-assured Air District data was downloaded from the EPA AirNow website from the Pleasanton and Livermore Air District monitoring stations. The Pleasanton $PM_{2.5}$ monitor is a Special Purpose Monitor (SPM) and a Federal Equivalent Method (FEM) monitor.[29] Specifically it is a FEM BAM 1020 monitor manufactured by Met One.[29] The Livermore $PM_{2.5}$ monitor, on the other hand, is a State or Local Air Monitoring Station (SLAMS) and a FEM monitor.[29] Like the Pleasanton station it is a FEM BAM 1020 manufactured by Met One.[29] Second, $PM_{2.5}$ data was downloaded from the publicly available low-cost sensor PA website using an API key provided along with Python code. PA sensors in the study area began recording data at different dates from 2018 to 2020 posing limitations for this analysis. **Figure 2** shows the PA sensors in the Tri-Valley available on September 28, 2020. Less than half of these stations were operating in 2018. This, along with the fact that PA sensors are deployed by users in the public are serious limitations in terms of the quality of data generated by these sensors. The PA data collected was _atm data for outdoor sensors and cf_1 data for indoor sensors. Cf_1 and cf_atm are "formulas used in Plantower laser counters" which are the mechanics that make up the PA sensors.[30] Cf_1 and cf_atm data have a proprietary correction factor formula automatically applied to it in order to compensate for different conditions outdoors versus indoors.[30] A previous study conducted by Barkjohn *et al.* (2022) used cf_1 data because it is more strongly correlated to reference monitors than cf_atm data.[31]

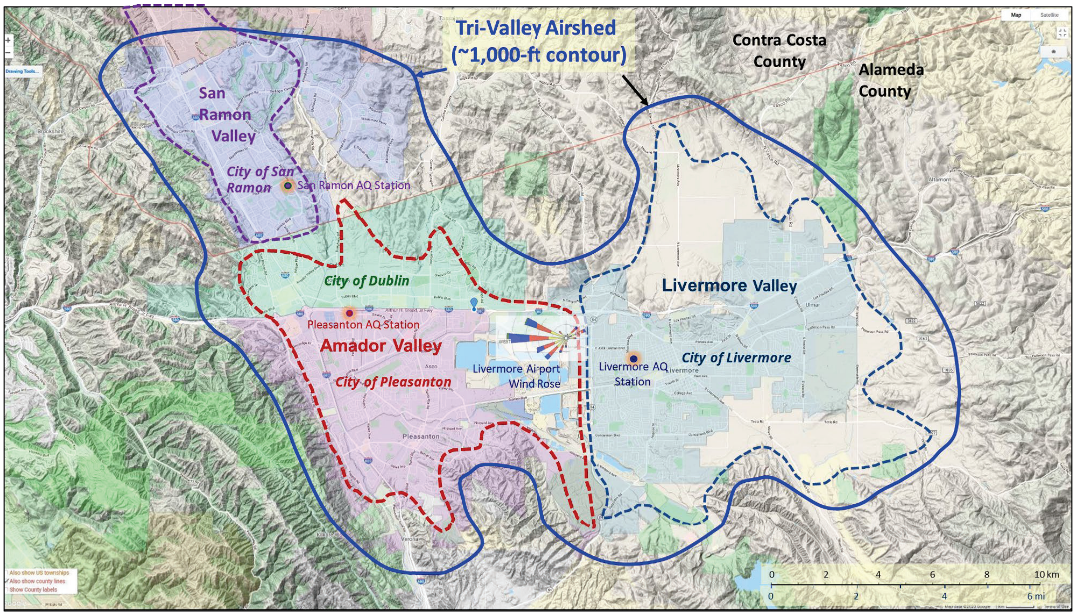

Figure 1. Tri-Valley Airshed in Contra Costa and Alameda counties in California, defined by the 1,000-ft contour (solid blue line) surrounding the San Ramon, Amador, and Livermore Valleys with valley floors depicted as dashed lines. The three BAAQMD Stations are noted as well as the annual wind rose at the Livermore Airport. (Created from Google Maps base)

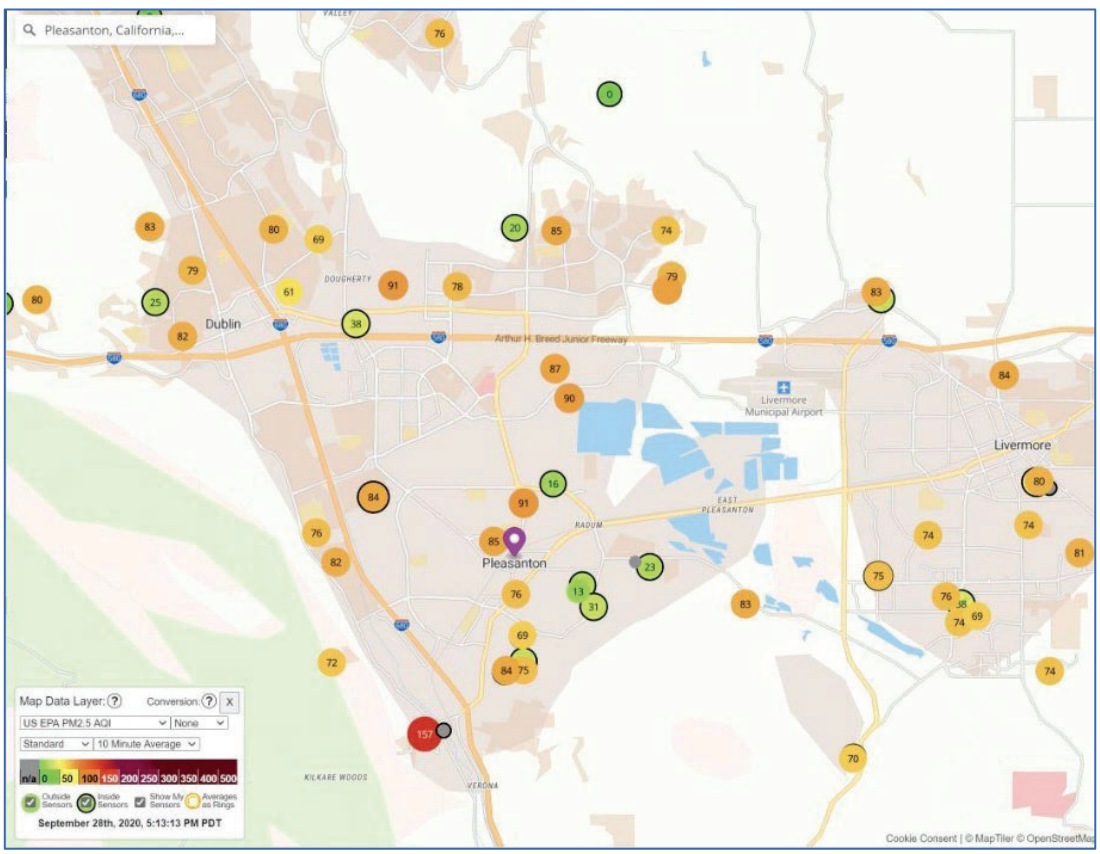

Figure 2. PurpleAir sensors in the Tri-Valley on September 28, 2020. Indoor stations are indicated with a solid black line around the station circle. For PurpleAir stations and their names used in this study, see **Figure 8.**

While accessing $PM_{2.5}$ data was relatively easy, gathering of hospital admissions data involved more effort and time (almost two months) and was a key learning experience of this study. The process began with reviewing previous wildfire health effect studies with County Health Department epidemiologists to select the relevant International Classification of Diseases 10 Clinical Modification (ICD 10 CM) codes that could distinguish smoke health effects. Since the Tri-Valley Airshed encompasses two counties, both Alameda County Public and Contra Costa County Health Departments were involved.

After several iterations, the final request involved epidemiologists from the two counties to write and execute scripts to download data from the California Department of Health Care Access and Information (HCAI) Patient Discharge Data and Emergency Department data:

- For asthma ICD 10 CM codes J45[2-5] [0-2], J459 plus COPD code J4[0-4] combined.
- Monthly by zip code each year from 2016 through 2020.
- By age groups <18, 18-44, 45-64, 65+, and total.

The expectation that sufficient daily visit data would be available for the four communities was not met. Instead, due to the small numbers and the requirement to maintain anonymity, monthly data were provided. Even with monthly totals, the number of visits occasionally was less than 11 and was denoted as <11 for anonymity.

Confirming periods when wildfire smoke was present in the Bay Area

Four tests were conducted to evaluate the difference between $PM_{2.5}$ levels during the 2018 and 2020 fire events and typical seasonal $PM_{2.5}$. The year 2019 was used as a control because based on reviewing daily NASA WorldView satellite imagery, no smoke plumes were found to enter the Bay Area that year. Other major wildfires during the study period include the September 2018 Walker Fire and the October 2019 Kincade Fire. However, NASA WorldView satellite imagery showed that the smoke from the Walker Fire located in the Sierra Mountains east of Chico was driven directly eastward and the plume from the Kincade Fire in Sonoma County went offshore to the west. The Bay Area escaped both plumes.

Four tests were run using Air District $PM_{2.5}$ data. The first test compares the Livermore Air District data from August 16, 2020, until October 8, 2020, to the same period in 2019. For the second test, the same time periods (August 16, 2020, through October 8, 2020, versus August 16, 2019, through October 8, 2019) are compared, but for the Pleasanton Air District sensor. The third test compares the Livermore sensor data from November 9, 2018, until November 27, 2018, to the same period in 2019. For the final test, the same November time periods are compared, but for the Pleasanton Air District sensor. The test periods were selected based on fires with the highest smoke impact from 2016 to 2020; years before 2016 were not included based on hospitalization and ED visit data availability.

Previous studies have used dispersion modeling to determine when the $PM_{2.5}$ increases were due to wildfire smoke plumes. In lieu of modeling the smoke trajectory, we initially selected days from known wildfires in the vicinity and narrowed the periods based on elevated $PM_{2.5}$ data. To independently select days when wildfire smoke was present over the Tri-Valley, we reviewed NASA WorldView satellite imagery. For the periods of interest, we viewed the daily Visible Infrared Imaging Radiometer Suite (VIIRS) Corrected Reflectance imagery taken by the VIIRS instrument aboard the joint NASA/NOAA NOAA-20 (JPSS-1) satellite. We included the VIIRS Fire and Thermal Anomalies layer which indicates active fire detections and thermal anomalies, such as volcanoes, and gas flares.

Figure 3 shows that while the Camp Fire was located over 150 miles from the Tri-Valley, strong Diablo winds from the northeast blew the smoke plume into the Tri-Valley Airshed. **Figure 4** plots the daily Air Quality Index (AQI) data taken in the Tri-Valley. (Note: An AQI of 100 represents 100% of the concentration of the associated federal standard.) **Figure 5** shows that the SCU Fires were much closer to the Tri-Valley. During the SCU Complex, the smoke plume enveloped the Tri-Valley Airshed with mean daily $PM_{2.5}$ AQI reaching 140 to 158 during three heavy smoke periods (**Figure 6**). On the first day of these three periods, ozone also was near or above the federal standard. Ozone has been found to be generated in wildfire plumes.[32] With its greater distance, ozone was only half the standard during the Camp Fire.

Figure 3. Satellite image of Camp Fire smoke on November 11, 2018. Source: NASA WorldView.

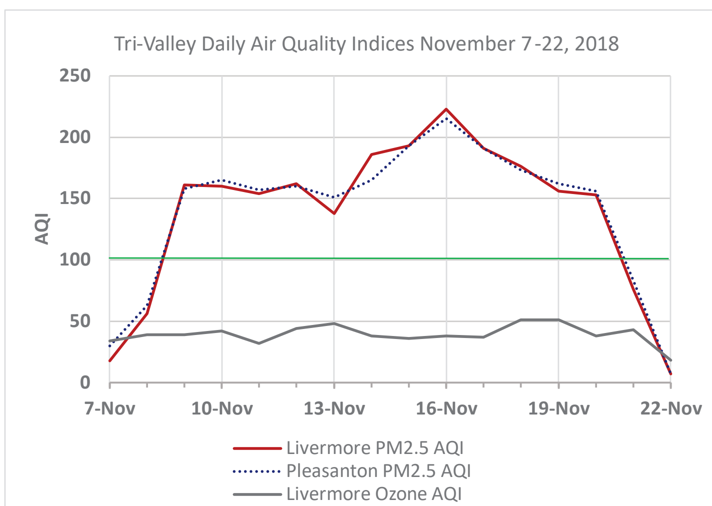

Figure 4. Daily ozone and $PM_{2.5}$ AQI at the Air District Pleasanton and Livermore stations from November 7-22, 2018. The green line (an AQI of 100) represents 100% of the concentration of the associated federal standard. Data Source: EPA AirData.

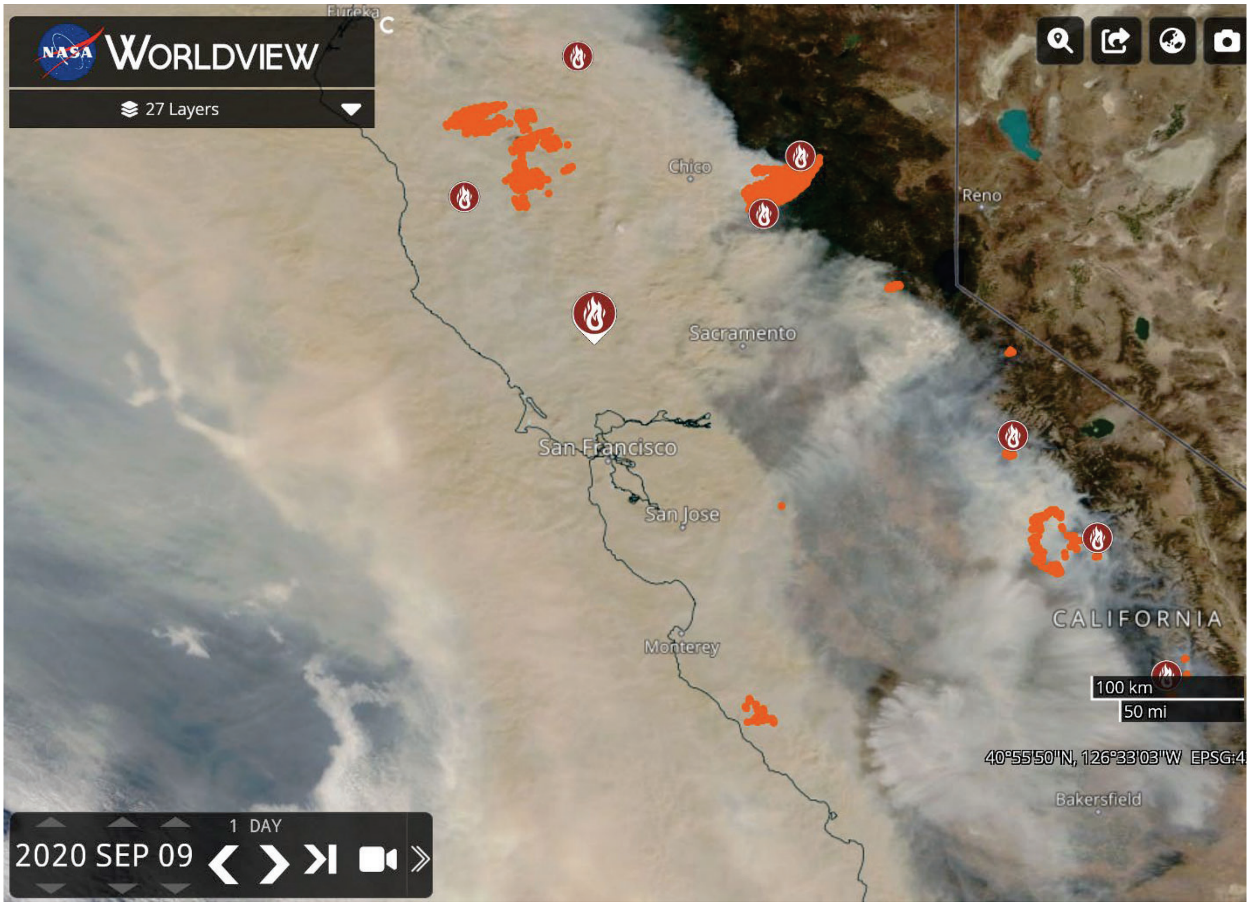

Figure 5. Satellite image of SCU Lightning Fire Complex on September 9, 2020. Source: NASA WorldView.

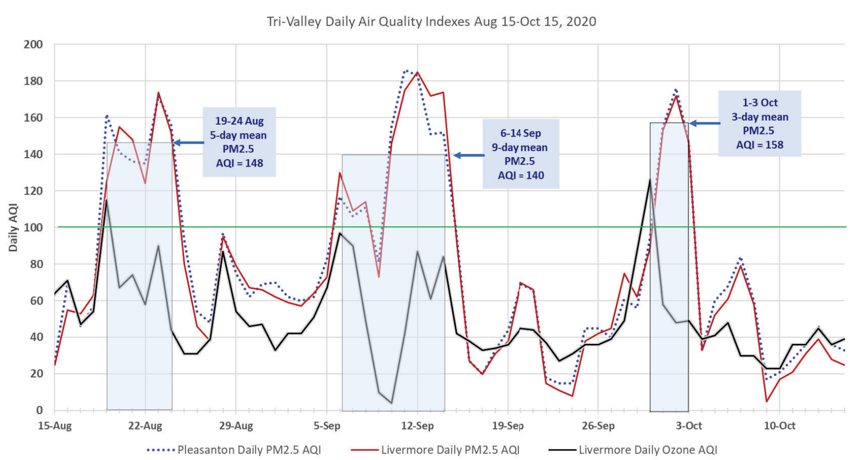

Figure 6. Daily ozone and $PM_{2.5}$ AQI at the Air District Pleasanton and Livermore stations from August 15 to October 15, 2020. The green line (an AQI of 100) represents 100% of the concentration of the associated federal standard. Data source: EPA AirData.

Correlation between PurpleAir and Air District sensors

PA sensors near to a quality-assured Air District monitor were compared to determine if a correlation exists. With this test we aim to understand if PA sensors in the Tri-Valley record at a similar precision and accuracy to Air District monitoring stations during periods where no wildfire smoke is present and during periods where wildfire smoke is present. However, this analysis requires some assumptions including that the district station and the PA sensor are under the similar conditions such as elevation, wind direction, as well as potential interference from surrounding manmade structures and that there are no other nearby point sources of $PM_{2.5}$ emissions. Sensor data from 2019 and 2020 were used for this comparison because more PA sensors were available closer to Air District stations than previous years. $PM_{2.5}$ data (both Air District and PA) was not collected past 2020, because respiratory hospitalizations and ED visits data was only available up until 2020. The closest outdoor sensor to the Livermore Air District station is the Valley Montessori sensor at 1.4 km (0.88 mi) away. The indoor sensor near the Livermore Air District station is the Heligan Lane PA sensor 2.6 km (1.62 mi) away. The outdoor PA sensor used for the Pleasanton Air District sensor is the James Dougherty Elementary School sensor 2.0 km (1.25 mi) away. And lastly, the indoor sensor used for the Pleasanton Air District sensor is the Dublin Mazda sensor at 0.5 km (0.33 mi) away. For a visual representation of sensor locations, please refer to **Figure 8**. Accuracy of the PA sensors was determined by linear regression, t-tests, and confidence intervals. These tests reveal if the average PA measurements are the same as the average Air District measurements, and if not, they reveal which sensor average is larger.

Geographical variation of $PM_{2.5}$ across the Tri-Valley during wildfire episodes

Another goal of the study was to determine how air quality varies across the Tri-Valley during wildfire episodes. To assess this, average $PM_{2.5}$ concentrations of 12 sensors across the Tri-Valley from August 16, 2020, to October 8, 2020, were computed.

Comparison of monthly average $PM_{2.5}$ with monthly hospitalizations data

This study's main goal was to quantify the association between monthly respiratory hospitalizations and $PM_{2.5}$ via linear regression analysis. We also stratified the analyses by age range. The original plan was to compare hospitalization and ED data with PA sensors for each ZIP code, however the hospital data at the ZIP code scale were not sufficient to perform this analysis. We explored using other respiratory conditions in addition to asthma and COPD, however the County Health Departments explained that using pneumonia and bronchitis would likely be too complicated due to noise from non-wildfire factors. Also, because respiratory hospitalizations in 2020 were almost all due to COVID-19, that year's data was not usable. The key limitations associated with the publicly-available hospitalization data were that only monthly totals were available, 2020 was not useable due to COVID-19, and hospitalizations of 11 or lower were recorded as <11 (for computations, values recorded as <11 were set to 11)

RESULTS

Hypothesis test and confidence intervals for periods when wildfire plumes were present versus absent

Overall, results were as expected—both the 2018 and 2020 fire period had higher average $PM_{2.5}$ concentrations than 2019 for both Livermore and Pleasanton (**Figure 7**.) The confidence intervals at 95% for the 2018 fire period versus the 2019 non-wildfire plume period, for both Livermore and Pleasanton, revealed that the wildfire influenced period mean $PM_{2.5}$ is greater than the non-fire period mean by 31.47 to 74.52 µg/m^3 and by 31.38 to 71.91 µg/m^3 respectively.

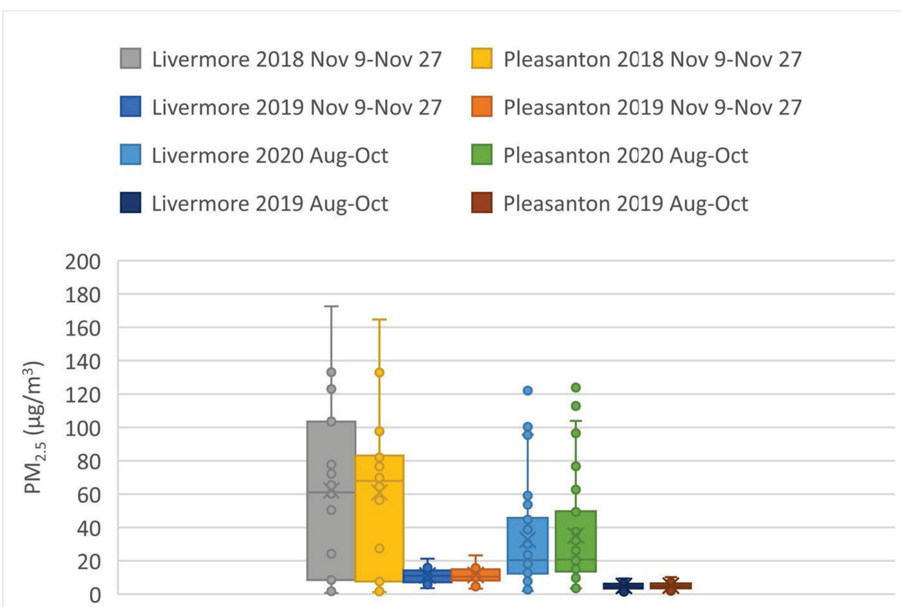

Figure 7. Comparison of the daily mean $PM_{2.5}$ concentrations at the two Air District stations (Livermore and Pleasanton) for the Camp Fire period, SCU Lightning Fire Complex period, and the associated time spans during 2019, when wildfire smoke was not present.

The 95% confidence intervals for the August through October 2020 fire period versus the 2019 non-fire period for both Livermore and Pleasanton revealed that the 2020 fire period mean $PM_{2.5}$ is greater than the 2019 non-fire period mean by 20.91 to 34.46 μg/m³ and by 22.5 to 37.57 μg/m³ respectively. In summary, $PM_{2.5}$ concentrations in Pleasanton and Livermore are on average significantly higher during the wildfire periods studied than the non-wildfire periods studied.

Correlation between PurpleAir and Air District sensors

Figure 8 shows the 12 PA outdoor sensors available in the Tri-Valley in 2019-2020 with their average concentrations during the 2020 SCU Lightning Fire Complex period. The closest outdoor and indoor PA sensor near each Air District sensor was selected to evaluate the accuracy of PurpleAir sensors. Note that different periods of record were available for each of the PA sensors. During the 2020 SCU Lightning Fire Complex period (August 16 to October 8, 2020), the average $PM_{2.5}$ concentration from the 12 sensors ranged from 30.64 to 51.15 μg/m³. The lowest $PM_{2.5}$ readings occurred at Valley Montessori School, Aspen Court, and Pasatiempo, likely due to sporadic missing data from August 16 through September 2, 2020. We were unable to conduct a similar analysis for the 2018 wildfire period because there was only one PA sensor recording at that time. Additionally, this sensor (Castlewood) had erroneous recordings, recordings that were outside the range of possible $PM_{2.5}$ recordings, such as readings over 1,000, and was therefore eliminated from the study. These erroneous values occurred both during periods when a wildfire smoke plume was present in 2020 and when one wasn't present in 2019.

Figure 8. The red dots show the 12 PurpleAir sensors used to evaluate $PM_{2.5}$ concentrations across the Tri-Valley. The size of the red dots represents the range in average $PM_{2.5}$ concentrations during the 2020 SCU Lightning Fire Complex period (August 16 to October 8, 2020). Note that the data used for this map is cf_1 data.

Sensor	Mean $PM_{2.5}$ (μg/m³)
Livermore Air District	32.57
Pleasanton Air District	35.16
James Dougherty	45.20
JM Amador	46.54
Dublin Onyx	49.80
Dublin Mazda (indoor)	33.93
Los Alamos	43.02
Lucca Circle	47.39
Mine*	47.78
Chateau PL*	38.75
Chardonnay	43.11
Pleasanton Hills	51.15
Heligan Lane (indoor)	11.67
Valley Montessori*	33.59
Aspen Court*	42.77
Pasatiempo*	30.64

Table 1. Mean $PM_{2.5}$ readings from Air District sensors, as well as PA sensors for the 2020 wildfire period (August 16, 2020 – October 8, 2020, except where noted with an asterisk). Sensors with an asterisk (*) had less data points (days with recordings) available and only began recording as early as September 1, 2020. Rows highlighted in grey represent the four PA stations that are compared to the Air District stations.

Air District (AD) Station	Nearest PurpleAir (PA) Station	Distance between Stations (km)	Coefficient of Determination (r^2)	Comparison of $PM_{2.5}$ Mean Values	95% Confidence Interval ($\mu g/m^3$)
Livermore	Valley Montessori (outdoor)	1.4	0.57	PA>AD	5.13 to 12.36
	Heligan Lane (indoor)	2.6	0.43	AD>PA	4.2 to 6.7
Pleasanton	James Dougherty (outdoor)	2.0	0.64	PA>AD	2.26 to 5.81
	Dublin Mazda (indoor)	0.5	0.79	AD>PA	3.35 to 5.57

Table 2. Correlation between Air District and PurpleAir $PM_{2.5}$.

Supplemental Figure 1 shows that the outdoor Valley Montessori PA sensor was moderately correlated to the nearby Livermore Air District sensor with a linear regression r-squared of 0.57. The confidence interval revealed at the 95% confident level the Valley Montessori PA sensor data is on average greater than the Air District sensor data by 5.13 to 12.36 $\mu g/m^3$. Delp *et al.* found a similar trend in PA II readings overpredicting $PM_{2.5}$ concentrations during the Camp Fire across Northern California.[33] Furthermore, these researchers found that an adjustment factor of 0.48 could make PA II readings more accurate for the Camp Fire in Northern California.[33] **Supplemental Figure 2** shows that the r-squared was reduced by 0.065 when only looking at the 7-week wildfire period compared to looking at the four-months from September to December 2020. Assuming the Livermore Air District readings are the accurate readings, the difference between the PA and Air District mean for each period gives a rough indicator for which period the PA readings are more accurate. During the entire period (September 2, 2020, to December 31, 2020) the mean difference between the Valley Montessori PA sensor and the Livermore Air District sensor was 8.7, whereas during the SCU Wildfire period the mean difference was 1.0, implying the Valley Montessori PA sensor was more accurate during just the SCU Wildfire period.

Both PA sensors in Pleasanton showed a good correlation with the Pleasanton Air District station. The James Dougherty Elementary outdoor sensor provided a reasonably good estimate of the Air District sensor with an r-squared of 0.64 (**Supplemental Figure 4**) We are 95% confident the true mean $PM_{2.5}$ of the James Dougherty sensor was greater than the Air District sensor by 2.26 to 5.81 $\mu g/m^3$. This suggests that the James Dougherty sensor recorded higher $PM_{2.5}$ levels than the Air District sensor on average despite their strong correlation. The Lane Regional Air Protection Agency found similar results when looking at PA sensors in their airshed in Oregon. They found that even though the PA sensors showed low accuracy to the reference monitors, they showed high precision with consistent overpredictions at double the reference monitor recordings.[34] When comparing **Supplemental Figure 4** and **Supplemental Figure 5**, the proportion of statistical variation of the James Dougherty Elementary PA sensor readings that can be explained by the Pleasanton Air District remains approximately the same between the SCU Wildfire period (August 16, 2020 to October 8, 2020) and the entire period studied (September 16, 2019, to December 31, 2020) This is supported by similar r-squared values of approximately 0.64 for both periods. Because the PA sensor mean is only approximately 4 points different than the Air District mean during the entire period, but 10 points different during the SCU Wildfire period, the PA sensor is likely less accurate during the SCU Wildfire.

Supplemental Figure 3 shows that the Heligan Lane indoor PA sensor was weakly correlated with the nearby Livermore Air District sensor with an r-squared of 0.426. The two-sample mean hypothesis test between the Heligan Lane indoor PA sensor, and the Livermore Air District sensor yielded a p-value of 4.76×10^{-13} which means there is statistically significant evidence that the two means are not the same. The confidence interval revealed at the 95% confident level the Air District sensor is on average 4.2 to 6.7 $\mu g/m^3$ greater than the Heligan Lane sensor. This is likely due to improved air quality indoors as the result of HVAC and air filtration systems.
On the contrary, **Supplemental Figure 6** illustrates that the indoor Dublin Mazda Pleasanton PA sensor showed strong correlation with an r-squared of 0.79, but the t-test showed the Air District sensor had higher concentrations on average. The t-test yielded a p-value of 2.69×10^{-11}. This t-test shows 95% confidence that the Pleasanton Air District sensor was on average greater than the Dublin Mazda sensor by 3.35 to 5.57 $\mu g/m^3$. The HVAC system plays a role in reducing indoor concentrations.

Comparison of monthly average $PM_{2.5}$ with monthly hospitalization data
The main takeaway from comparing $PM_{2.5}$ and Asthma and COPD hospitalizations is that there are serious limitations to statistical comparisons at this temporal scale. Monthly data going back only five years doesn't seem to be enough to draw

definitive conclusions with respect to the present study data, especially when the 2020 data was unusable due to the confounding effects of COVID-19.

Due to the 2020 hospitalization and ED data being affected by COVID-19, only the November 2018 fire period was available for analysis of wildfire effects. **Figure 9** shows the combined monthly average $PM_{2.5}$ from the two Air District sensors in Livermore and Pleasanton was not correlated with the monthly hospitalizations and ED visits for asthma and COPD for May through December 2018. This is likely because there is not enough data. If, for example, monthly November hospitalizations for many more years were used, better results might be possible to show whether or not respiratory ED and hospital visits are correlated with $PM_{2.5}$. Also, using individual months limits the variability associated with seasonality that occurs throughout the year. However, **Figure 10** reveals that both hospitalizations plus ED visits and $PM_{2.5}$ peaked in November 2018 during the Camp Fire. A simple line plot doesn't allow for a definitive correlation like a regression analysis does, but it does visualize the phenomena of hospitalizations/ED visits for respiratory issues increasing when $PM_{2.5}$ from wildfire smoke also increases.

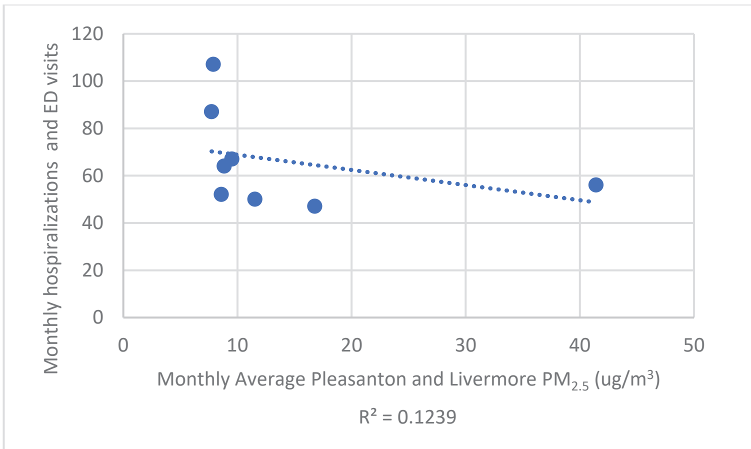

Figure 9. Monthly Average $PM_{2.5}$ from combined Air District sensors versus total monthly respiratory (asthma and COPD) hospitalizations and ED visits for May through December 2018.

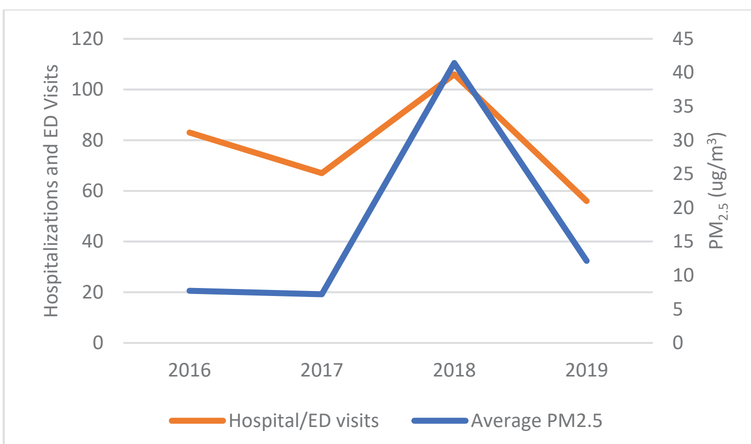

Figure 10. November Hospitalizations and ED visits plotted with average November $PM_{2.5}$ for 2016 through 2019. There is a notable spike in both $PM_{2.5}$ and respiratory hospitalizations and ED visits in November 2018 when the Camp Fire occurred.

DISCUSSION

Confirming periods when wildfire smoke was present in the Bay Area While the wildfires of 2018 and 2020 were located some distance from the Tri-Valley, hypothesis tests confirmed that dramatic $PM_{2.5}$ increases did occur in the Tri-Valley from these remote wildfires.

Correlation between PurpleAir and Air District sensors

The hypothesis tests and linear regressions between the PA sensors and Air District monitors confirmed that PA are moderately to strongly correlated with the Air District monitors for $PM_{2.5}$ recordings across the Tri-Valley. Based on previous literature, PA sensors overestimate $PM_{2.5}$ concentrations during periods of high concentrations, such as wildfires.[35, 31] Two recent studies found that the low-cost sensors were moderately to strongly correlated with the reference instrument, but that they overpredicted concentrations.[36, 33]

In this study, the linear regression between the outdoor Livermore PA sensor and the Livermore Air District sensor revealed a moderate correlation. However, it is interesting to note that the lower values seem to follow a linear trend, while the higher values do not. There was also a moderate relationship between the indoor Livermore PA sensor and the Air District sensor, with an r-squared value of 0.43. It was expected that the indoor PA sensor would be less correlated with the Air District sensor than the outdoor one.

The regression between the Pleasanton PA and the Pleasanton Air District outdoor sensors showed evidence for a strong linear relationship with an r-squared value of 0.64. It's interesting to note that the Air District sensor is located adjacent to the freeway, while the PA sensor, the James Dougherty Elementary sensor, is not located quite as close. It's also interesting that there is a strong correlation between the Air District and PA sensors in Pleasanton, but not necessarily for Livermore—especially when the Pleasanton James Dougherty sensor was farther away from its Air District sensor than the Livermore Valley Montessori sensor from its Air District sensor.

For both outdoor PA sensors, the mean difference between the PA sensor and the Air District sensor was compared for the entire period studied to just the SCU wildfire period to establish a rough indication of when the PA sensor is more accurate. There were conflicting results as the Livermore Valley Montessori PA sensor seems to be more accurate during the SCU Wildfire period and the Pleasanton James Dougherty PA sensor seems to be less accurate during the SCU Wildfire period. It was expected that both sensors would be less accurate during the SCU Wildfire period due to the higher concentrations of $PM_{2.5}$. However, there are confounding factors (such as a sensor that was misplaced/not working correctly, at one point, being fixed) that may be responsible for this discrepancy, along with the fact that comparing mean difference is only a rough indication for accuracy.

Another important finding is that the indoor Pleasanton sensor (Dublin Mazda) has a strong linear relationship with the Pleasanton Air District sensor with an r-squared value of 0.79. This is the strongest linear relationship of all four regressions conducted. This is likely because the Dublin Mazda sensor is so close to the Pleasanton Air District sensor (0.5 km apart), the closest out of all four sensors. Thus, the proximity of the sensors would explain the close correlation in the readings between the indoor PA sensor and outdoor Air District sensor, despite the fact one is indoors and the other is outdoors. Another plausible explanation for the strong linear relationship is that there is a lot of outdoor air flowing through the Dublin Mazda center. It could also be possible that the sensor was incorrectly labeled as indoor when it is actually located outdoors. A visit to the location could not confirm the sensor location.

Association between monthly average $PM_{2.5}$ and monthly hospitalization data

Because of the limitations to monthly hospital data, conclusions about the correlation between wildfire $PM_{2.5}$ and respiratory hospitalizations and ED visits in the Tri-Valley are not possible. Linear regression revealed no correlation between monthly $PM_{2.5}$ and hospital and ED visits. Previous studies were successful in quantifying the association between health effects and elevated $PM_{2.5}$ due to wildfires. For example, a study on the respiratory health effects of Northern California fires in 2008 found a "linear increase in risk for asthma hospitalizations… and asthma ED visits…with increasing $PM_{2.5}$ during the wildfires" with a risk ratio of 1.07.[2] Our current study's results don't match this trend due to our study area being much smaller (making the sample size even smaller), as well as only having access to monthly data and only for a five-year period.

CONCLUSIONS

Our study confirmed increases in $PM_{2.5}$ concentrations in the Tri-Valley during 2018 and 2020 wildfire episodes were statistically significant. When compared to BAAQMD sensors, low-cost PA sensors in the Tri-Valley were more accurate at lower concentrations, as previous studies had suggested. There was little variation in $PM_{2.5}$ across the Tri-Valley during the SCU wildfire episodes, but all sensors recorded average $PM_{2.5}$ concentrations above the federal standard daily average. Using regression analysis for monthly hospitalization, ED data, and monthly $PM_{2.5}$ concentrations is not adequate for evaluating wildfire effects on respiratory health. Daily hospitalization and air quality data are likely needed for future studies to quantify their relationship by regression analysis. These findings help inform future studies on air quality at community scales.

ACKNOWLEDGMENTS

This work was funded by a James Cary Smith Grant No. 2022.028 from the Bay Area Air Quality Management District. We give special thanks to Professor Ian Faloona at the University of California, Davis, for his valuable insights and guidance in the design of this research. His ideas and suggestions significantly enhanced the quality of this study. We greatly appreciate Matt Beyers, Roxanna Guide, and Elaine Bautista with the Community Assessment Planning & Evaluation (CAPE) Unit, Alameda County

Public Health Department, and Lisa Diemoz, and Jessica Cunningham-Krahl with Contra Costa Health Services for explaining and providing the respiratory hospitalizations and ED visit data.
Views and conclusions are those of the authors.

REFERENCES

1. Tri-Valley Air Quality Climate Alliance. (2023, November 1) *Tri-Valley airshed: What is it? https://tvaqca.org/?page_id=102*
2. 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*
3. Melvin, A. (2022) In *Wildland fires: Toward improved understanding and forecasting of air quality impacts*. Washington, DC; National Academies Press.
4. Baldassari, E. (2018, November 11) Camp fire death toll grows to 29, matching 1933 blaze as state's deadliest. *East Bay Times*.
5. U.S. Department of Commerce and Zimmerman, J. (2020) *Service Assessment, November 2018 Camp Fire*. Salt Lake City, Utah: National Weather Service.
6. Brewer, M. J., & Clements, C. B. (2019) The 2018 Camp Fire: Meteorological analysis using in situ observations and numerical simulations. *Atmosphere*, *11*(1), 47. *https://doi.org/10.3390/atmos11010047*
7. Borsum, D., & Plouffe, K. (2022, March 13) *SCU Lightning Complex, Central California*. ArcGIS StoryMaps. *https://storymaps.arcgis.com/stories/1ee955466a484df0950c28b57476bd65*
8. Walsh, J. (2020, August 22) Two major SCU lightning complex fires merge south of Tri-Valley. *Pleasanton Weekly*. Retrieved September 21, 2023, from *https://pleasantonweekly.com/news/2020/08/22/two-major-scu-lightning-complex-fires-merge-south-of-tri-valley*.
9. Kramer, A. L., Liu, J., Li, L., Connolly, R., Barbato, M., & Zhu, Y. (2023) Environmental justice analysis of wildfire-related $PM_{2.5}$ exposure using low-cost sensors in California. *The Science of the total environment*, *856*(Pt 2), 159218. *https://doi.org/10.1016/j.scitotenv.2022.159218*
10. O'Dell, K., Ford, B., Burkhardt, J., Magzamen, S., Anenberg, S. C., Bayham, J., Fischer, E. V., & Pierce, J. R. (2022)) Outside in: The relationship between indoor and outdoor particulate air quality during wildfire smoke events in western US cities. *Environmental Research: Health*, *1*(1), 015003. *https://doi.org/10.1088/2752-5309/ac7d69*
11. Environmental Protection Agency. (2024, January 22) *Managing Air Quality - Ambient Air Monitoring*. EPA. *https://www.epa.gov/air-quality-management-process/managing-air-quality-ambient-air-monitoring*
12. Environmental Protection Agency. (2024b, March 6) *National Ambient Air Quality Standards (NAAQS) for PM*. EPA. *https://www.epa.gov/pm-pollution/national-ambient-air-quality-standards-naaqs-pm*
13. Air Quality Sensor Performance Evaluation Center, Polidori, A., Papapostolo, V., Zhang, H., (2016, August) Laboratory Evaluation of Low-Cost Air Quality Sensors
14. Barkjohn, K. K., Gantt, B., & Clements, A. L. (2021) Development and application of a United States-wide correction for $PM_{2.5}$ Data collected with the PurpleAir sensor. *Atmospheric Measurement Techniques*, *14*(6), 4617–4637. *https://doi.org/10.5194/amt-14-4617-2021*
15. Wallace, L., Zhao, T., & Klepeis, N. E. (2022) Calibration of purpleair pa-I and PA-II monitors using daily mean PM2.5 concentrations measured in California, Washington, and Oregon from 2017 to 2021. *Sensors*, *22*(13), 4741. *https://doi.org/10.3390/s22134741*
16. Southerland, V. A., Brauer, M., Mohegh, A., Hammer, M. S., van Donkelaar, A., Martin, R. V., Apte, J. S., & Anenberg, S. C. (2022) Global Urban Temporal Trends in fine particulate matter (PM2·5) and attributable health burdens: Estimates from Global Datasets. *The Lancet Planetary Health*, *6*(2) *https://doi.org/10.1016/s2542-5196(21)00350-8*
17. Sun, Y., Mousavi, A., Masri, S., & Wu, J. (2022) Socioeconomic disparities of low-cost air quality sensors in California, 2017–2020. *American Journal of Public Health*, *112*(3), 434–442. *https://doi.org/10.2105/ajph.2021.306603*
18. *California Air Resources Board*. Ambient Air Monitoring - Regulatory | California Air Resources Board. (n.d.) *https://ww2.arb.ca.gov/our-work/programs/ambient-air-monitoring-regulatory*
19. Delfino, R. J., Brummel, S., Wu, J., Stern, H., Ostro, B., Lipsett, M., Winer, A., Street, D. H., Zhang, L., Tjoa, T., & Gillen, D. L. (2009) The relationship of respiratory and cardiovascular hospital admissions to the Southern California wildfires of 2003. *Occupational and Environmental Medicine*, *66*(3), 189–197. *https://doi.org/10.1136/oem.2008.041376*
20. Kiser, D., Metcalf, W. J., Elhanan, G., Schnieder, B., Schlauch, K., Joros, A., Petersen, C., & Grzymski, J. (2020) Particulate matter and emergency visits for asthma: A time-series study of their association in the presence and absence of wildfire smoke in Reno, Nevada, 2013-2018. *Environmental Health: A Global Science Source. https://doi.org/10.21203/rs.2.21450/v1*
21. Gan, R. W., Ford, B., Lassman, W., Pfister, G., Vaidyanathan, A., Fischer, E., Volckens, J., Pierce, J. R., & Magzamen, S. (2017) Comparison of wildfire smoke estimation methods and associations with cardiopulmonary-related hospital admissions. *GeoHealth*, *1*(3), 122–136. *https://doi.org/10.1002/2017gh000073*

22. Liu, J. C., Wilson, A., Mickley, L. J., Dominici, F., Ebisu, K., Wang, Y., Sulprizio, M. P., Peng, R. D., Yue, X., Son, J.-Y., Anderson, G. B., & 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*
23. Stowell, J. D., Geng, G., Saikawa, E., Chang, H. H., Fu, J., Yang, C.-E., Zhu, Q., Liu, Y., & Strickland, M. J. (2019) Associations of wildfire smoke PM2.5 exposure with cardiorespiratory events in Colorado 2011–2014. *Environment International*, *133*, 105151. *https://doi.org/10.1016/j.envint.2019.105151*
24. Leibel, S. (2019) Increase in pediatric respiratory visits associated with wildfires in San Diego County. *Journal of Allergy and Clinical Immunology*, *143*(2) *https://doi.org/10.1016/j.jaci.2018.12.071*
25. Aguilera, R., Corringham, T., Gershunov, A., & Benmarhnia, T. (2021) Wildfire smoke impacts respiratory health more than fine particles from other sources: Observational evidence from Southern California. *Nature Communications*, *12*(1) *https://doi.org/10.1038/s41467-021-21708-0*
26. Alexeeff, S. E., Deosaransingh, K., Van Den Eeden, S., Schwartz, J., Liao, N. S., & Sidney, S. (2023) Association of long-term exposure to particulate air pollution with cardiovascular events in California. *JAMA Network Open*, *6*(2) *https://doi.org/10.1001/jamanetworkopen.2023.0561*
27. Teresa Feo. (2023, September) *The Human Health Benefits of Improving Forest Health in California: Investigating the Links Between Forest Management, Wildfire Smoke, and the Health Sector.* California Council on Science and Technology and Blue Forest. Sacramento, CA. *https://ccst.us/reports/ the-human-health-benefits-of-improving-forest-health-in-california/*
28. 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*, 186–195. *https://doi.org/10.1016/j.etap.2017.08.022*
29. Bay Area Air Quality Management District. (2023, June) *2023 Annual Air Monitoring Network Plan. https://www.baaqmd.gov/~/media/files/technical-services/2023_network_plan-pdf.pdf?rev=8de9f6f74a2143a994734a3a870bd999&sc_lang=en*
30. PurpleAir. (2023, September 13) *What is the difference between CF=1, ATM, and alt?*. PurpleAir Community. *https://community.purpleair.com/t/what-is-the-difference-between-cf-1-atm-and-alt/6442*
31. Barkjohn, K. K., Holder, A. L., Frederick, S. G., & Clements, A. L. (2022) Correction and accuracy of PurpleAir PM2.5 measurements for extreme wildfire smoke. *Sensors*, *22*(24), 9669. *https://doi.org/10.3390/s22249669*
32. Faloona, Ian C. Sen Chiao, Arthur J. Eiserloh, Raul J. Alvarez II, Guillaume Kirgis, Andrew O. Langford, Christoph J. Senff, Dani Caputi, Arthur Hu, Laura T. Iraci, Emma L. Yates, Josette E. Marrero, Ju-Mee Ryoo, Stephen Conley, Saffet Tanrikulu, Jin Xu, and Toshihiro Kuwayama. (2020) The California Baseline Ozone Transport Study (CABOTS) *Bulletin American Meteorological Society.* April 2020. *https://doi.org/10.1175/BAMS-D-18-0302.1*
33. Delp, William W. and Brett C. Singer. (2020) Wildfire Smoke Adjustment Factors for Low-Cost and Professional PM2.5 Monitors with Optical Sensors. *Sensors* 2020, 20, 3683; doi:10.3390/s20133683
34. LRAPA. (2018) *LRAPA PurpleAir Monitor Correction Factor History.* LRAPA; Springfield, OR, USA
35. Johnson, K., A. Holder, S. Frederick, and A. Clements. (2020) PurpleAir PM2.5 U.S. Correction and Performance During Smoke Events 4/2020. International Smoke Symposium, Raleigh, NC, April 20-24, 2020. *https://cfpub.epa.gov/si/si_public_file_download.cfm?p_download_id=540979&Lab=CEMM*
36. Holder, A. L., Mebust, A. K., Maghran, L. A., McGown, M. R., Stewart, K. E., Vallano, D. M., Elleman, R. A., & Baker, K. R. (2020) Field evaluation of low-cost particulate matter sensors for measuring wildfire smoke. *Sensors*, *20*(17), 4796. *https://doi.org/10.3390/s20174796*

ABOUT STUDENT AUTHOR

Natasha Atkins completed her Bachelor of Science in Environmental Science and Management at the University of California, Davis, in June 2023. She plans to continue her education in environmental sciences in the near future.

PRESS SUMMARY

In the summer and fall of 2018 and 2020, major wildfires in Northern California impacted the San Francisco Bay Area. This study was designed to analyze the increase in health effects caused by wildfire smoke using available hospital data at the community level in the Tri-Valley area of the East Bay Area. However, since only monthly hospital data were available publicly from California databases, the relationships could only be implied. Daily hospitalization data are required to quantify their relationship by regression analysis. This study revealed that low-cost PurpleAir sensors likely gave more accurate concentrations during typical air quality but likely tended to overestimate respirable particle air concentrations when wildfire plumes were present.