ASSESSING BIAS IN PERSONAL EXPOSURE ESTIMATES WHEN INDOOR AIR QUALITY IS IGNORED: A CASE STUDY OF EASTERN NORTH CAROLINA AND RALEIGH/DURHAM AREA

By

Abdulahi Opejin

July 2024

Director of Thesis: Misun Hur, PhD
Major Department: Geography, Planning, and Environment

Abstract

Exposure to outdoor air pollution can cause adverse health issues. Indoor air pollution can be more dangerous due to indoor activities such as cooking, painting, smoking, and heating systems, and can worsen the burden associated with air pollution. Neglecting indoor air quality in an exposure assessment may generate bias and erroneous conclusions regarding the association between exposure and health outcomes. Accurately assessing levels of exposure to air pollution helps advance our understanding of the health effects associated with air pollution and identify effective strategies to address public health concerns. Previous exposure studies have utilized outdoor air quality to assess individuals' exposure to air pollution in their residential locations. These studies neglect the significant amount of time people spend indoors and other activity places. Some other studies assessed exposure to air pollution by measured only indoor air quality inside homes, but they overlook outdoor and other indoor environments where people are exposed to air pollution. Therefore, it remains unknown whether neglecting indoor air quality would generate negligible or non-negligible bias. Likewise, environmental justice studies have also focused only on outdoor air quality to understand exposure disparities among social groups. It remains unknown whether the degree of disparities in exposure becomes larger when

indoor environments are considered compared to when only outdoor air quality is considered in exposure assessments.

This study assessed these biases by comparing two types of personal exposure estimates for 100 individuals: one derived from real-time particulate matter (PM2.5) measurements collected both indoors and outdoors using a low-cost portable air monitor (GeoAir2.0) and the other from PurpleAir sensor network data collected exclusively outdoors. It also investigates whether exposure disparity between different social groups would be more prominent when indoor air quality is considered in exposure assessments. The PurpleAir measurement data were used to create smooth air pollution surfaces using geostatistical methods. To obtain mobility-based exposure estimates, both sets of air pollution data were combined with the individuals' GPS tracking data. Paired-sample t-tests were then performed to examine the differences between these two estimates. This study also investigated whether GeoAir2.0- and PurpleAir-based estimates yielded consistent conclusions about gender and economic disparities in exposure by performing Welch's t-tests and ANOVAs and comparing their t-values and F-values. The study revealed significant discrepancies between GeoAir2.0- and PurpleAir-based estimates, with PurpleAir data consistently overestimating exposure ($t = 5.94$; $p < 0.001$). It also found that females displayed a higher average exposure than males (15.65 versus. 8.55 μg/m3) according to GeoAir2.0 data ($t = 4.654$; $p = 0.055$), potentially due to greater time spent indoors engaging in pollution-generating activities traditionally associated with females, such as cooking. This contrasted with the PurpleAir data, which indicated higher exposure for males (43.78 versus. 46.26 μg/m3) ($t = 3.793$; $p = 0.821$). Additionally, GeoAir2.0 data revealed significant economic disparities ($F = 7.512$; $p < 0.002$), with lower-income groups experiencing higher exposure—a disparity not captured by PurpleAir data ($F = 0.756$; $p < 0.474$). These findings highlight the importance of considering both indoor and outdoor air quality to reduce bias in exposure estimates and more accurately represent environmental disparities.

ASSESSING BIAS IN PERSONAL EXPOSURE ESTIMATES WHEN INDOOR AIR QUALITY IS IGNORED: A CASE STUDY OF EASTERN NORTH CAROLINA AND RALEIGH/DURHAM AREA

A Thesis

Presented to the Department of Geography, Planning, and Environment faculty.

East Carolina University

A Partial Fulfilment of the Requirements for the Degree

Master of Science in Geography

By

Abdulahi Opejin

July 2024

Director of Thesis: Misun Hur, PhD

Thesis Committees:

Yoo Min Park, PhD

Sinan Sousan, PhD

Arthur Liu, PhD

ACKNOWLEDGMENTS

First, I want to express my profound gratitude to my primary (informal) advisor, Dr. Yoo Min Park, for her persistent guidance and efforts, which have been instrumental since the beginning of my master’s program in helping me achieve my academic and research goals. Without her, this research would not have been successful.

I want to thank my thesis committee chair and formal advisor, Dr. Misun Hur, and committee members, Drs. Sinan Sousan and Arthur Liu, for their critical roles and invaluable contributions to my research.

I want to express my gratitude to the Department of Geography, Planning, and Environment for constantly supporting me through summer funding for data collection and their support throughout the program. My appreciation would be incomplete without a special thanks to Drs. Yoo Min Park and Sinan Sousan for allowing me to use the data from their pilot project funded by the National Institute of Environmental Health Sciences of the National Institutes of Health under Award No. P30ES025128 through the Center for Human Health and the Environment, North Carolina State University.

I sincerely appreciate the other people who helped me in one way or another through this research and my academic journey: my parents, brothers, sisters, friends, and other faculty members, Mr. Scott Wade and Dr. Mukherji, for their unwavering support and motivation throughout my program.

TABLE OF CONTENTS

List of Tables

List of Figures

1 CHAPTER ONE: INTRODUCTION

1.1 Background of the study

Exposure to air pollution is known to cause adverse health outcomes, such as asthma, heart disease, stroke, chronic obstructive pulmonary disease, and lung cancer (Buonanno et al., 2014; Steinle,2015; Klepeis et al., 2017; WHO, 2022). Human activities, such as the combustion of fossil fuels for motor vehicles, industrial processes, and power generation, release substantial amounts of gases and particulates into the air and contribute significantly to outdoor air quality (Kampa & Castanas, 2008). Indoor activities, including cooking, painting, cleaning products, combustion of candles, and heating systems, aggravate the burden of health risks associated with air pollution (Mocumbi et al., 2019; Omelekhina et al., 2020; Vardoulakis et al., 2020). Therefore, it is essential to consider both indoor and outdoor sources of air pollution to accurately assess personal exposure and its health impacts (Dias & Tchepel, 2018; Susanne et al., 2015; Vu et al., 2022). Failure to adequately capture the spatiotemporal variations in air quality across various indoor and outdoor environments may potentially underestimate or overestimate the actual levels of exposure to pollutants (Wang et al., 2023; Yoo et al., 2021). The bias may generate erroneous conclusions regarding the association between exposure and health outcomes. This bias could also make it challenging to understand the unequal burden of air pollution among different social groups. Therefore, improving the accuracy of exposure assessments is essential as it can enhance our understanding of the effects of air pollutants on human health and overall well-being, particularly for marginalized populations, such as groups with low socioeconomic status (Caplin et al., 2019; Kim & Kwan, 2021).

Researchers have used various air pollution data and methods for decades to estimate people's exposure levels. Many previous studies have combined outdoor air pollution measurement data or air pollution modeling outcomes with individuals' residential locations to estimate personal exposure levels (Bell, 2006; Breen et al., 2015; Son et al., 2010). However, the findings of these studies are influenced by the uncertain geographic context problem (UGCoP)—the problem that causes inferential errors and bias in health risk assessments when human mobility patterns are not considered in exposure assessments (Kwan, 2012). Since people's mobility is cut across different spaces and times, using the residential neighborhood as a contextual unit does not represent an actual geographic context in which people are exposed to air pollution. In recent years, there have been increased concerns about addressing the UGCoP to improve the accuracy of exposure assessments. As a result, several studies have evaluated human exposure to air pollution by combining outdoor air pollution data and people's mobility data (Chen et al., 2018; Lu, 2021; Nyhan et al., 2016; Park & Kwan., 2017; Yu et al., 2023). This method is a step forward because it considers human mobility in exposure estimation and partially addresses the UGCoP by going beyond residential neighborhoods. However, these studies did not consider exposure at indoor locations where people spend a significant amount of time. Other studies have assessed exposure levels using indoor measurements by placing low-cost air pollution sensors in fixed indoor locations, such as homes and offices (Doris et al., 2024; Dwivedi et al., 2024; Tong & Ho, 2018; Vicente et al., 2024). However, these studies focused merely on indoor exposure in these fixed locations, excluding other indoor and outdoor environments where people spend their time. Therefore, their research findings may still be affected by the UGCoP.

More recent studies have incorporated real-time personal air sampling data collected both indoors and outdoors using portable, low-cost air sensors and human mobility data to assess levels of exposure (e.g., Beko et al., 2015; Ma et al., 2020; Park et al., 2023; Steinle et al., 2014). They have emphasized that assessing exposure to environmental influences should not be limited to using data from an outdoor monitoring network and residential data because people move through various indoor and outdoor microenvironments to undertake daily activities. These studies contributed to methodological advances by distinguishing between indoor and outdoor environments in exposure assessments using mobility-based air sample data. For example, studies (Lu, 2021; Ma et al., 2020; Rao et al., 2023; Song et al., 2023) have compared air quality data from governmental monitoring stations with real-time mobility-based $PM_{2.5}$ concentrations from low-cost mobile air sensors. The studies found that assessments based on stationary outdoor monitoring data overestimated individual exposure levels when ambient air pollution levels were relatively high. However, the result of these studies might have been substantially affected by several factors: a mismatch in the temporal resolution of the two data sets (i.e., hourly vs. real-time), too few sample measurements available from the regulatory monitoring stations, and significant differences in accuracy between the reference monitors and low-cost sensors. To date, no studies have evaluated the discrepancy between mobile and stationary monitoring data when both types of data are collected at the same temporal scale, using low-cost sensors that have comparable accuracy and reasonable spatial resolutions. Additionally, previous studies focused mostly on urban areas where outdoor air quality is generally poor, and comparison research focusing on the United States' rural communities where air quality is moderate does not exist. Therefore, it is unknown whether ignoring indoor exposure may result in negligible or non-negligible bias in exposure estimates of people in rural areas in the United States, and if so, how much bias is introduced.

Addressing bias in exposure assessments is also an essential step in achieving environmental justice (EJ) as it enables researchers to more accurately identify the most vulnerable groups and more effectively tailor policy recommendations. Traditional EJ research has examined how differential social settings lead to differences in environmental exposures and health outcomes. The role of indoor air quality in understanding EJ is crucial because it introduces additional complexities; indoor air quality can be influenced by both indoor and outdoor sources, which may often be determined by various socioeconomic factors (Admkiewicz et al., 2011; Jing et al., 2018; Yu et al., 2024). However, most EJ studies have historically focused on the unequal distribution of outdoor sources of air pollution, such as industrial facilities or traffic emissions (Adamkiewicz et al., 2011). Indoor environments and human mobility have not been fully considered in EJ research. As a result, the potential for greater disparities in exposure, once indoor environments are accounted for alongside outdoor air quality in assessments, remains underexplored. New research is needed to capture the degree of disparity in exposure between different social groups more accurately by considering both indoor and outdoor air quality as well as human mobility.

In line with the above discussion, this study underscores the importance of expanding exposure assessment research to include all indoor and outdoor spaces where people encounter air pollution and experience disparities in exposure. Therefore, this study attempts to answer the following two research questions:

1. How much bias may be induced in a personal exposure assessment when only outdoor air quality data is integrated but indoor air quality is ignored?
2. Does real-time mobile monitoring data, which include measurements from both indoor and outdoor environments, reveal larger differences in exposure levels among different social groups compared to data collected solely from outdoor monitors?

In order to empirically answer the study's research questions, this study aims:

1. To assess bias in personal exposure by comparing estimates derived from two different air pollution data: (1) real-time personal air sampling data collected both indoors and outdoors using a low-cost, location-aware, portable air monitor (GeoAir2.0 [Park et al., 2021]) and (2) real-time outdoor concentration data from the distributed PurpleAir low-cost sensor network.

2. To investigate whether exposure disparities between different social groups would be more prominent when indoor air quality is considered in exposure assessments.

The remainder of this thesis is structured as follows: Chapter Two reviews the existing body of literature on air pollution exposure assessments, highlighting the key methodological limitations of previous exposure studies and the necessity for assessing errors in personal exposure estimates and revealing a hidden portion of environmental injustice due to the errors. Chapter Three describes the research methodology, including data collection, data manipulation, statistical analysis, and geospatial analysis for this study. Chapter Four presents the research findings, including results regarding the degree of bias in personal exposure estimates and assessments of gender/economic disparities in exposure. Chapter Five summarizes the major findings and discusses the significance of this research, contributions to the literature, and future considerations.

2 CHAPTER TWO: LITERATURE REVIEW

The literature review discusses the limitations of exposure assessment methods frequently used in the literature by critically reviewing previous studies that fail to consider both human mobility and highly localized air pollution data measured both indoors and outdoors. Therefore, it calls for empirical research that assesses bias in personal exposure estimates when human mobility and indoor air quality are ignored by incorporating data on individuals' mobility patterns and highly localized air pollution measurements from both indoor and outdoor settings.

2.1 Methodological limitations of previous exposure studies

Early studies estimated levels of exposure to air pollution using residential locations and outdoor air pollution data, assuming people were stationary and only exposed to outdoor air pollution in their residential neighborhoods (Figure 2.1). Several studies (e.g., Bell, 2006; Son et al., 2010) discovered that using only regulatory monitoring data to estimate exposure levels was inadequate, particularly for rural populations, because estimating rural residents' exposure based on monitors located in urban areas is challenging. Estimates at locations that are far from monitoring stations can have significant errors. Conventional air quality monitoring stations have limited spatial coverage, leading to significant errors in the estimation of $PM_{2.5}$ concentrations across large geographic areas. Bell (2006) used eight ozone monitoring stations to investigate the use of air quality modeling to estimate participants' exposure levels in the Northern Georgia Region of the United States. Son et al. (2010) utilized 13 regulatory monitoring stations to estimate human exposure levels across the study area based on residential locations. The research revealed that the central estimates of the relationship between air pollution and lung function might be impacted by the exposure estimation technique. The research concluded that spatial interpolation techniques like kriging could provide more accurate exposure level estimations than only monitoring data because they found that central health impact estimates from exposure estimations based on kriging or averages across all monitors were often greater than those from the closest monitor. This technique can capture the spatial variability of individual-level exposures and can generate estimates for areas lacking monitoring stations. One of the significant issues in the exposure assessment studies that focuses on only outdoor air quality is the unavailability of high-resolution air quality data because the existing outdoor monitoring stations are too dispersed to capture the hyper-local fluctuations in air quality.

Figure 2.1 Estimation of pollutant exposure in residential homes

More recent studies have begun to use spatiotemporally higher-resolution air pollution data collected from low-cost air sensors. Low-cost air sensors have offered numerous opportunities for building a denser monitoring network to capture high-resolution air quality data. Masiol et al. (2018) utilized low-cost Speck monitors to measure particulate matter (PM) concentrations at 48 outdoor sites near homes with wood-burning appliances. This data was then used to predict hourly PM concentrations and assess people's exposure levels at residential locations. The study developed hourly land-use regression (LUR) models to predict the hourly levels of $PM_{2.5}$ in the study area. To explain the geographical variability in $PM_{2.5}$ concentrations, the LUR models integrated land use information, such as road networks, parks, and industrial sites. These models examined the relationship between transient variations in air pollution and health consequences. The study revealed that deploying low-cost PM monitors at 23 sites significantly enhanced the spatial and temporal resolution of air quality data compared to traditional monitoring methods. This approach enabled the collection of detailed and comprehensive pollutant concentration data across the metropolitan area, capturing variations that could be missed using fewer monitoring station data.

Similarly, Chen and Lin (2022) combined monitoring station and low-cost sensor data to estimate daily average exposure using the inverse distance weighting method. This study highlights that the governmental monitoring station data offers limited spatial coverage, and focusing only on this data can lead to significant errors in spatial interpolation and $PM_{2.5}$ exposure estimation for health risk assessments. The study suggested that utilizing data from low-cost microsensors with smart spatial interpolation methods can be a valuable approach to improving exposure assessments. However, these studies only focused on outdoor air quality, neglecting the significant amount of time people spend indoors and indoor air pollution sources that contribute to their exposure levels. They also ignored human mobility and various daily activity locations in exposure assessments. Ignoring these factors may lead to inferential errors and inaccurate exposure assessments because people can also be exposed to air pollution in other places they visit. Therefore, people's mobility should be considered to assess exposure levels more accurately (Kwan, 2013).

For accurate assessments, several studies have utilized people's mobility data to estimate personal exposure levels at different locations they visit. Park and Kwan (2017) utilized outdoor air quality data from the national monitoring sites to estimate hourly air pollution levels using co-kriging (a spatial interpolation method) and then integrated them with individuals' simulated travel pattern data to assess individual exposure. The study demonstrated that assessing environmental exposure and related health risks might be significantly erroneous if the spatiotemporal dynamics of air pollution and human movement are ignored. By combining different kinds of data, including individuals' residential locations, simulated individual-level movement patterns, daily average pollution levels, and hourly air pollution levels, they performed three-dimensional (3D) geovisualization and geospatial technologies to illustrate how humans and air pollutants interact dynamically over space and time. In addition, they revealed how four categories of exposure estimates derived from a different combination of these data sets exhibited significant variations, highlighting the significance of taking both human mobility and spatiotemporal variability of air pollution into account. Similarly, Lu (2021) used hourly PM data obtained from a low-cost PurpleAir sensor network, where the travel patterns of individuals that were estimated using the South California Association of Government activity-based model data (SCAG-ABM) were used to predict static and dynamic exposure using a random forest model. The study found that static exposure estimates were less accurate for workers than for nonworkers and that they often underestimated actual exposure by roughly 13%. The bias in exposure estimation was particularly pronounced for people with higher mobility. These studies consider human mobility data, which is an improvement over the studies that focus only on residential locations. However, they rely only on outdoor air quality data. Research that focuses only on outdoor air quality, even though it considers human mobility, could still generate biased findings.

Recent studies have suggested that indoor air quality could be worse than outdoor air quality due to various sources of air pollution indoors. Because most people spend significant of their time indoors (approximately 90%), health risks associated with indoor air pollution can be higher than those related to outdoor exposure, which has become a growing public health concern (Omelekhina et al., 2020; Park et al., 2022; Streuber et al., 2022). Noticing the importance of understanding indoor exposure, Tong and Ho (2018) assessed personal exposure to indoor air pollution in 55 elderly households during winter and summer by placing low-cost monitors one meter above the ground in the living room of participants' homes. The study highlights the significance of improving both indoor and outdoor air quality to protect vulnerable groups from negative effects of air pollution and also revealed that approximately 50% of households exceeded the World Health Organization (WHO) indoor air quality standards for $PM_{2.5}$, with higher non-compliance in winter (61.0%) compared to summer (40.6%). Additionally, the study revealed that incense burning and cooking were significant sources of indoor $PM_{2.5}$ and NO_2. In a similar study, Klepeis et al. (2017) investigated the exposure level of low-income children to fine airborne particles in their homes by placing low-cost monitors three feet above the floor and one foot from the nearest wall and conducted interviews to gather information on home characteristics, ventilation, and particle sources. The study revealed no significant relationship between lower particle counts and ventilation activities. However, particle counts were associated with indoor smoking of marijuana or cigarettes, lighting of candles, house cleaning, and cooking.

These studies significantly contributed to advancing our knowledge of indoor air quality and its health impacts. However, they focused solely on home environments, thereby introducing the UGCoP in the exposure estimates. This approach neglects the outdoor environments and other indoor locations where people spend significant amounts of time, leading to incomplete and potentially inaccurate exposure assessments. To accurately assess personal exposure to air pollution, it is crucial to utilize high-resolution air quality data measured in both indoor and outdoor environments, combined with detailed human mobility data. This comprehensive approach allows for a more precise and realistic evaluation of individual exposure to air pollutants, capturing the full range of environments encountered in daily life (Figure 2.2).

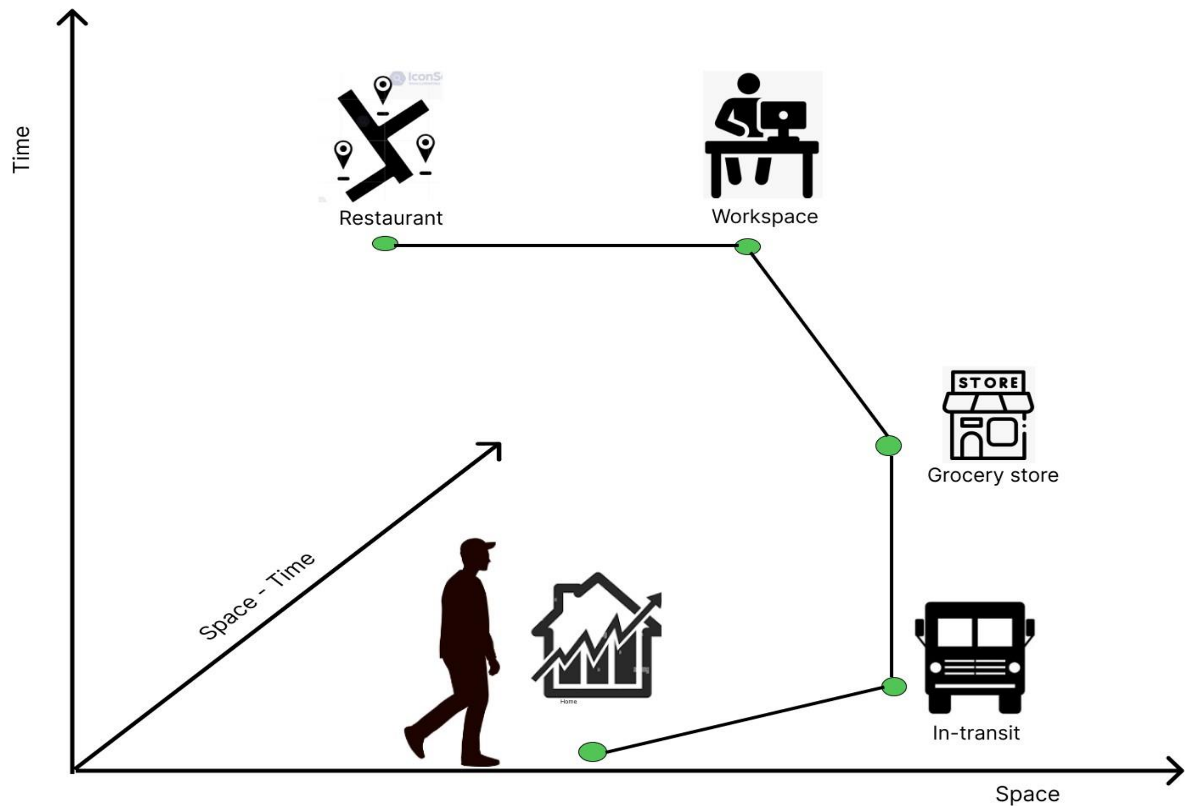

Figure 2.2 Individual's space-time movement patterns

2.2 The necessity for assessing errors in personal exposure estimates.

More recent studies that assess personal exposure to air pollution have acknowledged the importance of high-resolution real-time air quality data and human mobility data in exposure assessments. A few recent studies have incorporated real-time personal air sampling data collected both indoors and outdoors using portable, low-cost air sensors and human mobility data to assess levels of exposure. Steinle et al. (2014) used a portable monitor and a global positioning system (GPS) device to collect real-time air quality data and human mobility data from 17 research participants. The GPS and time-activity diary was used to track the pathways of people's daily activities in various indoor and outdoor environments. The study found significant variability in $PM_{2.5}$ concentrations across different microenvironments and activity places and, therefore, highlighted the importance of capturing comprehensive exposure pathways to assess individual exposure risks accurately. Beko et al. (2015) assessed the contribution of different microenvironments to personal exposure from sixty non-smokers using portable monitors and GPS tracking devices to collect particle number concentration data and people's location data, respectively. The study revealed that a significant fraction (50%) of daily personal exposure to ultrafine particles occurred within the home and about 90% of daily exposure occurred indoors, revealing the importance of indoor environments in overall exposure levels. These studies have highlighted the importance of incorporating both indoor and outdoor environments into personal exposure assessments. Therefore, it is necessary to understand how much bias can be introduced in exposure and health risk assessments when indoor exposure is not considered. However, there are very few empirical studies that examine the biases.

As an exception, Ma et al. (2020) compared different types of exposure estimates derived from residence-based assessments using 35 governmental monitoring station data, mobility-based assessments using monitoring station data, and mobility-based assessments using real-time air quality data from low-cost portable monitors. The study found that both mobility-based and residence-based assessments using outdoor monitoring station data cannot account for variations in exposure between indoor and outdoor environments. They also found that assessments based on stationary outdoor monitoring data overestimated individual exposure levels when ambient air pollution levels were relatively high and could generate bias in personal exposure estimation. However, the result of this study might have been substantially affected by several factors: a mismatch in the temporal resolution of the two data sets (i.e., hourly vs. real-time), too few sample measurements available from the governmental monitoring stations, and significant differences in accuracy between the reference monitors and low-cost sensors. To date, no studies have evaluated the discrepancy between mobile and stationary monitoring data when both types of data are collected at the same fine temporal scale, using low-cost sensors that have comparable accuracy and with reasonable spatial resolutions. Also, their study was conducted in Beijing, China, where outdoor air quality is generally poor, but comparison research focusing on the United States' rural communities does not exist. Therefore, it is unknown whether ignoring indoor exposure and people's movement patterns would result in non-negligible bias in exposure estimates of people in rural areas in the United States, and if so, how much bias may be introduced.

2.3 Underreported environmental injustice due to errors in exposure estimates

Uneven distribution of polluting sources and planning regulations that pose disadvantages to low-income and/or minority communities have led to an unequal distribution of risk. Therefore, addressing environmental injustice is important to mitigate health disparities. Research on environmental justice and health disparities has primarily focused on outdoor air pollution; thus, indoor air quality disparities have often been neglected. Addressing disparities in indoor air pollution is challenging due to the influence of both indoor and outdoor sources, which can be influenced by socioeconomic factors, such as housing quality and neighborhood environments (Adamkiewicz et al., 2011). Brown et al. (2015) investigated the relationship between indoor air quality and socioeconomic status, analyzing data from 567 households to examine disparities in indoor exposure. They found that lower-income households had higher formaldehyde concentrations, which can have negative health effects. The presence of smoke in indoor homes was also associated with increased levels of $PM_{2.5}$ due to frequent cooking, the presence of at least one smoker, and the frequent use of air fresheners. $PM_{2.5}$ levels are significantly higher in households with higher occupancy and low-income households. Similarly, Abdel-Salam et al. (2022) studied indoor PM_{10} levels in 60 domestic kitchens and found that households with lower income and education levels had higher levels of indoor pollutants. Adamkiewicz et al., (2011) integrated indoor environmental quality into the environmental justice (EJ) dialogue by identifying the individual and place-based drivers of indoor environmental exposures. The study highlighted that households with lower socioeconomic status (SES) tend to have higher concentrations of indoor pollutants, influenced by a combination of outdoor and indoor sources, building structures, and resident activities. All of these studies suggest a correlation between poor indoor air quality and lower socioeconomic status, highlighting the need to address environmental injustice and health disparities in indoor air pollution.

These studies improved an understanding of the unequal burden of indoor air pollution. However, what remains unknown is whether differences in exposure to air pollution between different income groups would become greater when both indoor and outdoor air quality data are considered, compared to when only outdoor air quality data is used. If the between-group differences increase when indoor air quality is incorporated into the analysis, it may suggest that traditional EJ research might have underestimated the degree of disparities in environmental exposure.

This literature review highlights the importance of considering both indoor and outdoor air quality as well as human mobility to mitigate the UGCoP and produce more reliable findings in environmental exposure research. It also argues that if between-group differences are smaller when only outdoor air quality is considered compared to when both outdoor and indoor air quality are incorporated, it may signify that the social disparity in environmental exposure has been underreported in previous EJ studies that focus on outdoor air. To support these arguments, this study compares the group differences estimated using real-time personal air sample data with those calculated using outdoor monitoring station data.

3 CHAPTER: METHODOLOGY

Real-time, geo-referenced personal air sampling data from GeoAir2.0 portable monitors, outdoor air pollution monitoring data from the PurpleAir low-cost sensor network, and participants' demographic data were used to assess potential bias in exposure estimates when human mobility and indoor air quality are ignored and assess whether disparities in exposure to air pollution are underreported.

3.1 Study area

The study area for this research is eastern North Carolina (ENC) and the Raleigh/Durham area (Figure 3.1). The annual average of outdoor $PM_{2.5}$ ranges between 12.1 - 59 µg/m³ in North Carolina, which appears moderate (North Carolina Air Quality Index, 2023). However, higher air pollutant concentration episodes may be experienced indoors due to different human activities such as cooking, cleaning, painting, etc., and could potentially cause events of peak exposure indoors, especially in low-income and socially marginalized populations. Eastern North Carolina is a region characterized by a heterogeneous population of individuals belonging to diverse ethnic groups with a proportional representation of residents with low- and high-income levels. The eastern North Carolina's average poverty rate of 17.6 % significantly surpasses both the North Carolina state average of 12.6 % and the national average of 13.3 % (Scharer, 2001).

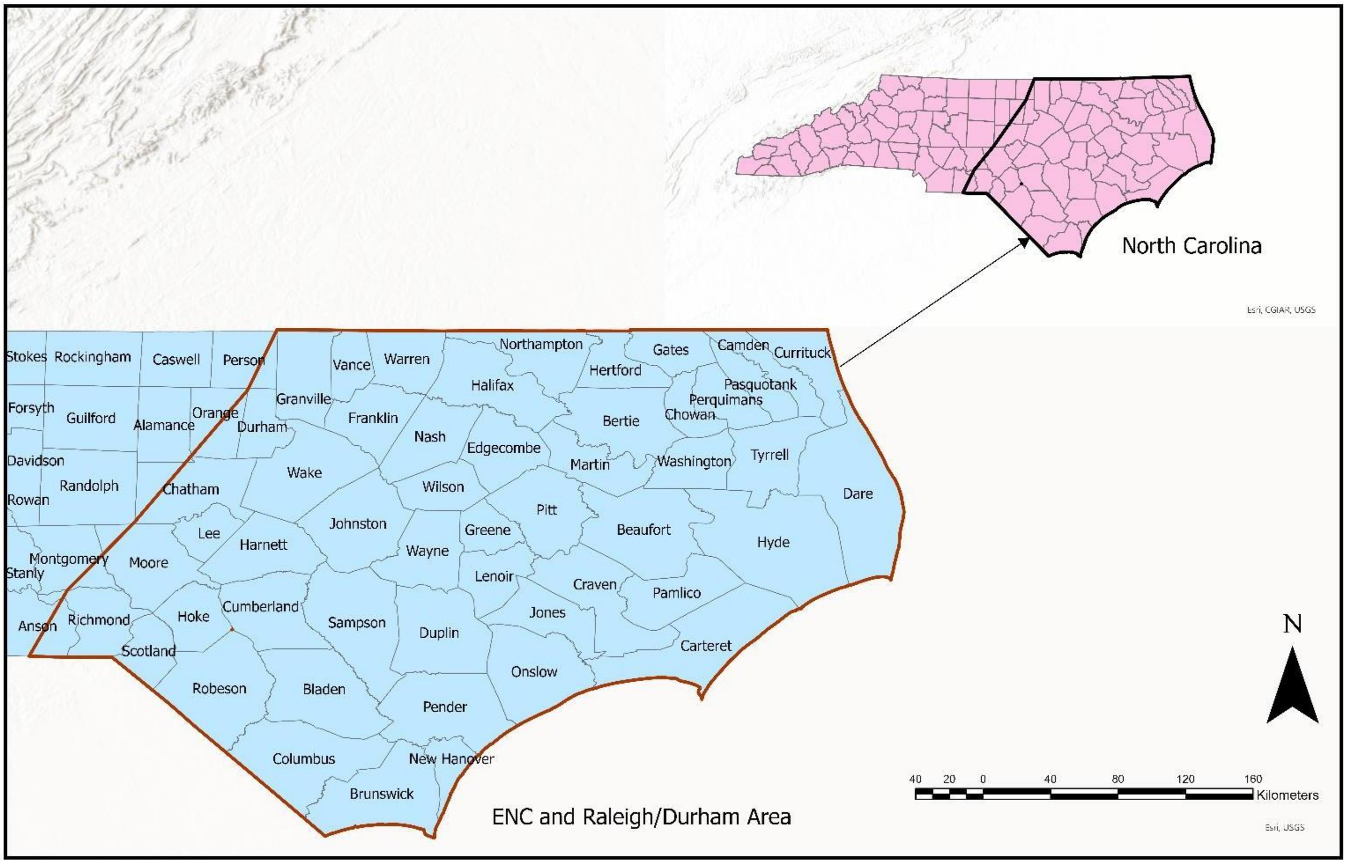

Figure 3.1 Study area: Eastern North Carolina and Raleigh/Durham Area

3.2 Data

3.2.1 Indoor and outdoor air pollution data from GeoAir2.0 portable air monitors

3.2.1.1 The specification and performance of GeoAir2.0 monitors

This study used GeoAir2.0 data collected from a previous pilot project conducted by Park's research team between October 2021 and March 2022. The methodology for participant recruitment and data collection is detailed in Park et al. (2023) paper. The GeoAir2.0, a low-cost air monitor, can collect personal air sampling data in various indoor and outdoor environments. It consists of a low-cost $PM_{2.5}$ sensor (SPS30, Sensirion, Switzerland), a global positioning system (GPS) module, a humidity/temperature sensor, a battery that can last about 15 hours, a liquid crystal display (LCD) screen, and a data logger (Park et al., 2021). The monitor can be worn with a carabiner or belt clip or trapped in the shoulders (Figure 3.2).

Various laboratory and field tests have confirmed its accuracy and scientific reliability when evaluated with reference instruments (Park et al., 2021; Sousan et al., 2021; Streuber et al., 2022). Streuber et al. (2022) conducted an extensive analysis to compare the performance of GeoAir2.0 to various commercially available air monitors. The study evaluated the accuracy of GeoAir2.0 for salt and dust aerosols and examined the effect of changes in relative humidity on its measurements using the MiniWRAS reference instrument in laboratory settings. The performance of GeoAir2.0 was also examined in indoor environments by comparing it to the reference instrument (pDR-1500) over five days in three homes. The GeoAir2.0 consistently demonstrates a strong correlation (r = 0.96–0.99) with the MiniWRAS for both salt and Arizona Road Dust (ARD) measurements in laboratory experiments. The GeoAir2.0 also correlated well with the filter-corrected pDR-1500 in residential indoor settings, with correlation coefficients ranging from 0.80 to 0.99. Also, the study revealed that changes in humidity had less of an impact on GeoAir2.0 than on a higher-cost sensor, OPC-N3. In particular, GeoAir2.0 indicated an increase in mass concentrations ranging from 100% to 137% for low and high concentrations. The accuracy of the SPS30 PM sensor, which is used in the GeoAir2.0 device, was also assessed by Sousan et al. (2021). The authors evaluated the performance of four commercially available, low-cost particulate matter (PM) sensors (SPS30, OPC-N3, AirBeam2, and PMS A003) in measuring PM mass concentrations in occupational and environmental settings. The sensors were calibrated by performing side-by-side measurements with a reference instrument using three aerosols: salt, Arizona road dust, and Poly-alpha-olefin-4 oil. In environmental settings, SPS30 revealed a very high correlation (r = 0.99) with the reference instrument for all aerosol types. In occupational settings, SPS30 maintained a high correlation ($r > 0.96$) with the reference instrument. These previous studies suggest that GeoAir2.0, which uses the SPS30 sensor, is effective in estimating PM mass concentrations in residential indoor, occupational, and environmental settings by producing data of a quality comparable to that of more expensive instruments.

Figure 3.2 A GPS-enabled low-cost portable air monitor (GeoAir2.0) and the three ways to wear it

3.2.1.2 Description of GeoAir2.0 data

The GeoAir2.0 simultaneously records $PM_{2.5}$ readings (µg/m^3), GPS positions, temperature, humidity, date, time, and other data at a 1-minute interval. It also collects a list of nearby Wi-Fi media access control (MAC) addresses to ensure indoor geolocation accuracy. The GeoAir2.0 data that this study utilized includes location data as well as place types for each GPS location.

Park et al. (2023) developed a data fusion method to (1) restore and rectify inaccurate GPS locations, particularly indoor locations, by combining GPS and Wi-Fi MAC address data provided by GeoAir2.0 and (2) automatically categorize GPS locations into homes, workplaces/schools, in-transit, and other places. While the initial study by Park et al. (2023) included 44 Latino/Hispanic individuals, this study used a larger set of data, including those obtained from the second-round data collection, which resulted in 56 additional participants from diverse race/ethnic groups. Consequently, this study involved a total of 100 participants living in eastern North Carolina (ENC), who were asked to carry GeoAir2.0 monitors for three days (i.e., two weekdays and one weekend day). Because the participants completed the data collection on different sets of three consecutive days, the data for this study included 100 participants' geo-referenced air pollution data for 90 unique calendar dates. To minimize selection bias in participant recruitment, Park's research team pilot survey employed various recruitment methods with only one inclusion criterion and no exclusion criteria. Specifically, anyone between 18 and 64 years old was eligible for participation in the study. Community partners' networks, social media platforms, word-of-mouth, and a door-knocking method were utilized to promote the research and ensure broad outreach and engagement with the target audience. Additionally, several local churches and businesses were visited, and short presentations about the project were given at a non-profit community organization, and A Better Chance A Better Community (ABC2) was specifically used to recruit Black participants. The University and Medical Center Institutional Review Board at East Carolina University reviewed and approved the study material and protocol on 7/25/2023 (IRB#: UMCIRB 23-001348). This approval indicates that the research protocol, ethical considerations, methodology, and participant consent procedures have complied with the ethical standards set by East North Carolina University IRB.

3.2.2 Outdoor air pollution data from a PurpleAir sensor network

3.2.2.1 The specification and performance of PurpleAir monitors

Although the data accuracy of low-cost air monitors is relatively lower than that of regulatory monitoring stations in general, low-cost sensors allow for the acquisition of data from numerous locations at a community level and produce air quality data with higher spatiotemporal resolutions than those from a regulatory monitoring network. Therefore, for the outdoor air pollution data, this study used real-time $PM_{2.5}$ data from a monitoring network that consisted of 213 PurpleAir low-cost sensors across ENC and its surrounding areas. The PurpleAir monitor uses a low-cost PMS5003 sensor (Plantower, Bejing, China) to detect PM ranging from 0.3 to 10µm, and the control board estimates the total mass for PM_1, $PM_{2.5}$, and PM_{10} using particle counts calculated every second. The data is transferred to the cloud every 80 seconds. This low-cost sensor network has been established through a crowd-sourced community effort to monitor air pollution concentrations at a fine spatiotemporal resolution to fill the data gaps in locations where regulatory monitoring stations are scarce (Figure 3.3). The PurpleAir monitors have been managed and maintained by community members, and the real-time air quality measurements have been provided through a publicly available web map (PurpleAir, 2023).

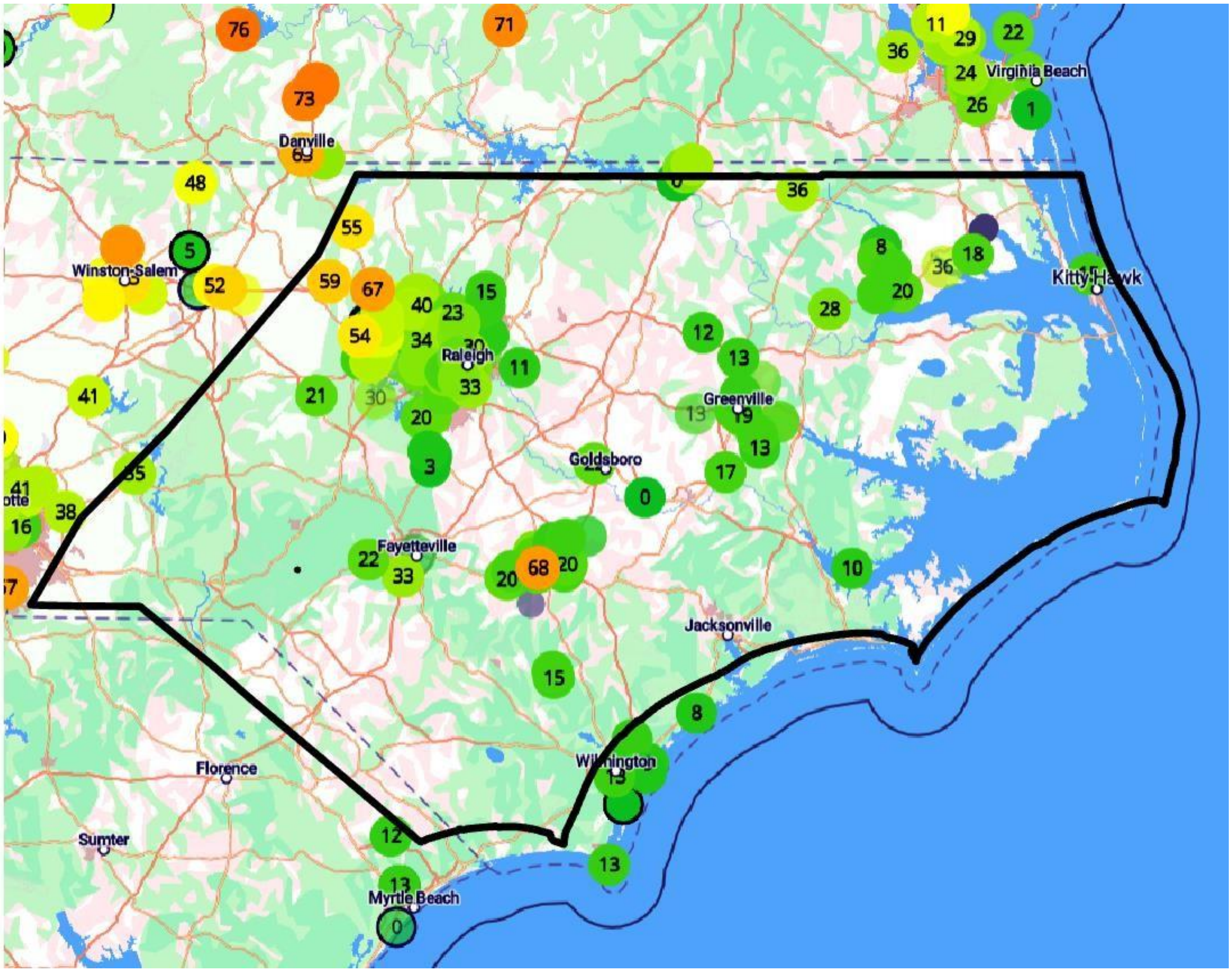

Figure 3.3 Distribution of PurpleAir monitors across the Eastern North Carolina and Raleigh/Durham Area (the PurpleAir web map)

Numerous studies have conducted laboratory and field tests to confirm the accuracy and scientific reliability of PurpleAir low-cost sensors by comparing them with reference instruments. For example, Stavroulas et al. (2021) compared the performance of PurpleAir sensors with established $PM_{2.5}$ measurement instruments, including Beta Attenuation Monitors (BAM) and the optical reference-grade monitor in two urban environments (Athens and Ioannina, Greece) across three seasons. The study revealed that PurpleAir correlates strongly with reference instruments with (R^2 = 0.87) for the Beta attenuation monitor and (R^2 =0.98) for the optical reference-grade monitor. The study highlights that reliable data for air quality monitoring can be obtained using PurpleAir sensors with appropriate calibration.

Jiang et al. (2021) conducted an analysis to evaluate the performance of the Plantower PM sensor with two established reference instruments - the gravimetric method (GM) and Tapered element oscillating microbalances (TEOM), which is the federal reference method (FRM) and federal equivalent method (FEM) respectively, both approved by US Environmental Protection Agency (EPA). The study evaluated the accuracy of the PurpleAir sensor using laboratory experiments with $PM_{2.5}$ concentrations ranging from 8 to 167 μg/m³ and assessed the consistency of the sensors in both laboratory and field settings. The performance of the PurpleAir sensors was also examined in various environments by comparing it to the reference instruments (TEOM and GM) during field evaluations. The study revealed that PurpleAir correlated well with the reference instruments in laboratory and field settings by consistently demonstrating strong correlations ($r \geq 0.82$) with relatively small relative standard deviations (16–21%) between sensors., confirming its potential applicability for $PM_{2.5}$ exposure assessment. After calibration, the PurpleAir sensor measurements became more convergent with the reference methods, indicating that linear calibration methods may be sufficient for ambient monitoring using TEOM. In contrast, nonlinear calibration methods may be more appropriate for indoor air quality monitoring using GM. This highlights that the PurpleAir sensor can produce data of comparable quality to more expensive, high-precision instruments.

3.2.2.2 Description of PurpleAir data

The PurpleAir data used in this study includes outdoor $PM_{2.5}$ concentration measurements in every minute, dates, times, and geographic coordinates of the monitor locations. It was downloaded using the PurpleAir API and PurpleAir data download tools (Opejin, 2024a). To ensure temporal alignment between the PurpleAir and GeoAir2.0 data, this study extracted timestamps in the GeoAir2.0 data and used it to download PurpleAir data to ensure that the two data are uniform and correspond to the same times and dates.

3.2.3 Demographic Survey

This study used the survey of socioeconomic data collected from the pilot project mentioned above. The survey includes information about participants' socioeconomic and demographic characteristics, including gender, age, race/ethnicity, and household income levels. Physical addresses of participants' homes and workplaces/schools were also collected. For all questions, the option to "prefer not to respond" was provided to allow participants to avoid sharing the information if they preferred. The data was used to examine bias in assessments of economic and gender disparities in exposure.

3.3 Data Analysis

3.3.1 Geospatial Data Analysis

Because the GeoAir2.0 data was already processed by Park et al. (2023), it did not require additional data processing for this study and was ready to use. To generate air pollution surfaces and calculate personal exposure using PurpleAir data, this study developed a Python algorithm using geospatial packages, including ArcPy and Geopandas, to perform the geostatistical analysis (ordinary kriging) and data extractions efficiently (Opejin, 2024b). The algorithm systematically filtered the PurpleAir dataset using the timestamp field to select instances of PurpleAir monitors available for a particular minute. The $PM_{2.5}$ values from the selected PurpleAir monitors were used to perform a kriging operation to generate an air pollution surface at a specific minute on a particular date (Figure 3.4). The air pollution surface was then combined with participants' GPS trajectory data (i.e., latitude and longitude) available in the GeoAir2.0 dataset to calculate PurpleAir-based personal exposure estimates. Participants' GPS trajectory data was dynamically overlaid on the kriging output to extract each participant's air quality values for that corresponding minute. Because the PurpleAir dataset had 129,600 unique minutes, which corresponds to the number of unique minutes in the GeoAir2.0 dataset, the process was iterated 129,600 times. All the geoprocessing and data integration procedures were completed using Pandas and Matplotlib Python packages.

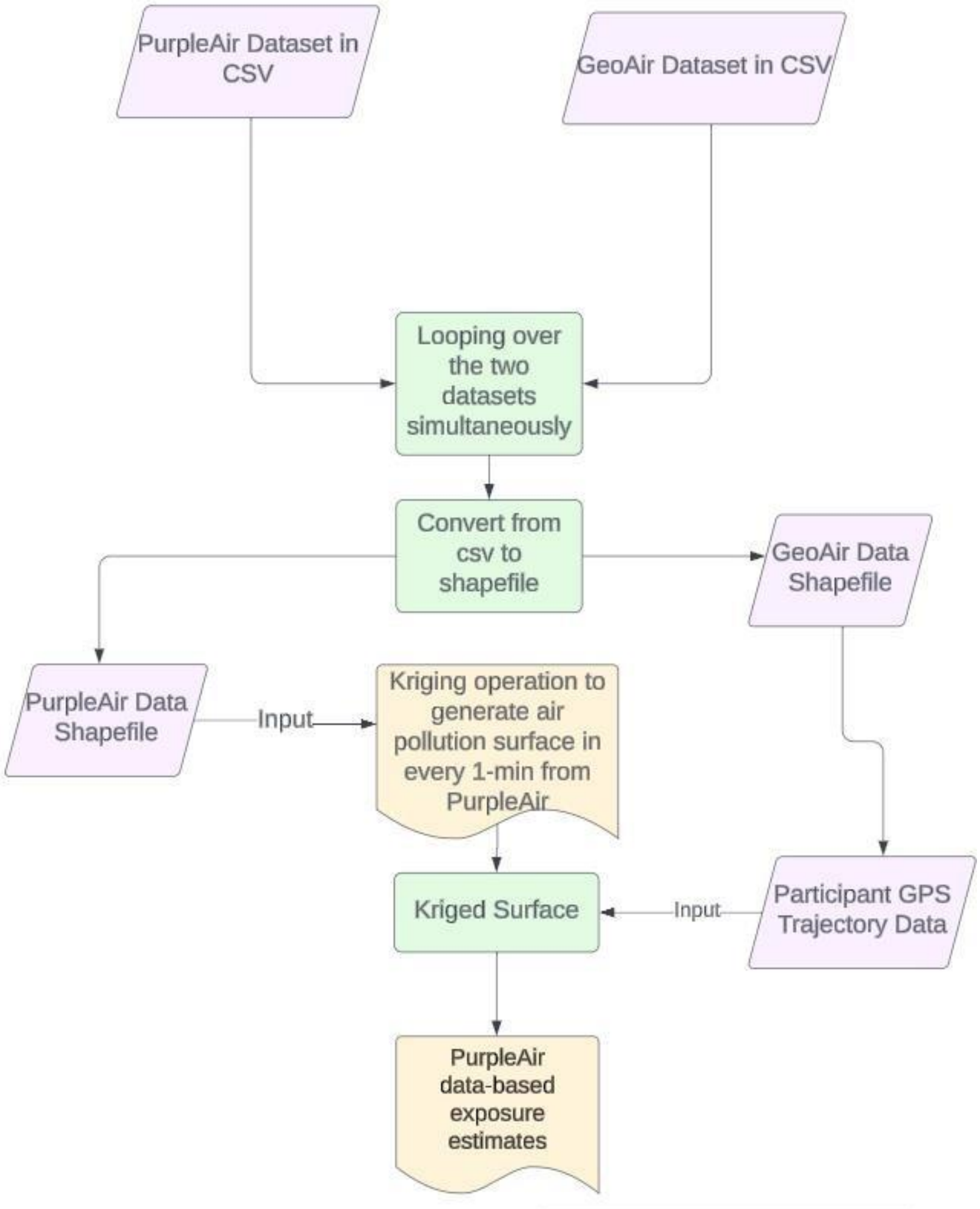

Figure 3.4 Flowchart of geospatial data analysis procedures to estimate participants' exposure levels using PurpleAir data

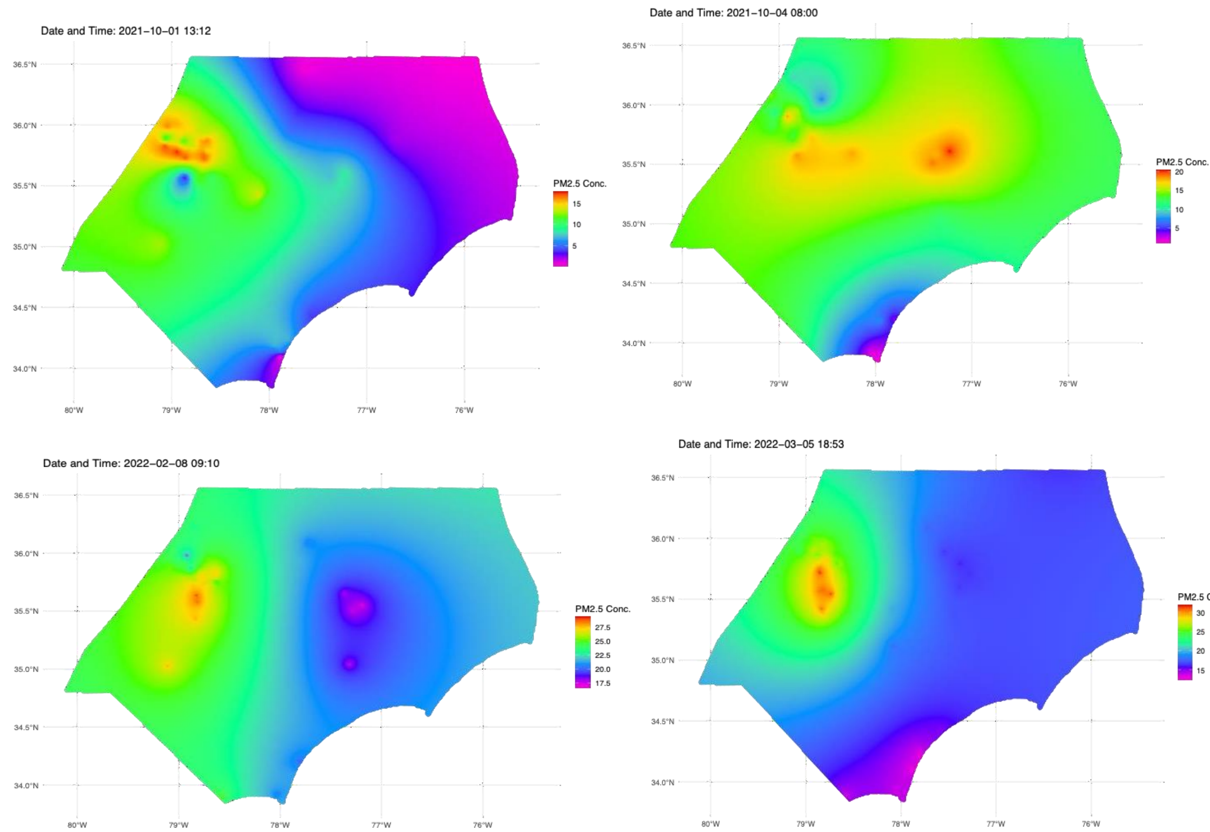

Figure 3.5. Examples of kriging outcomes based on the PurpleAir low-cost sensor network data (1-minute intervals from September 2021 to March 2022)

3.3.2 Statistical Data Analysis

3.3.2.1 Bias in levels of exposure

The study examined whether the two exposure estimates (i.e., GeoAir2.0- and PurpleAir-based estimates) are significantly different. Participants' $PM_{2.5}$ exposure values were averaged for each dataset to obtain total exposure, weekday vs. weekend exposure, and exposure by different activity locations (e.g., homes, workplaces, in transit, and other places). Then, paired-sample t-tests were performed to examine whether there was a significant difference between the two estimates. Likewise, for home exposure, the two datasets were filtered to retain only home exposure, and then paired-sample t-tests were performed to examine whether there was a significant difference in home exposure between the two exposure estimates. The same procedure was utilized to assess the discrepancy between the two estimates for exposure during transit, at workplaces, at other places, and during weekdays and weekends.

3.3.2.2 Bias in the magnitude of social disparities in exposure

To assess the extent of economic disparities in exposure revealed by each set of air pollution data, the participants were categorized into three income groups: 37 participants with income less than $30,000 as low-income, 25 with income between $30,001 and $75,000 as middle-income, and 38 with income more than $75,000 as high-income. The Welch's analysis of variance (ANOVA) was used for each type of data to assess whether significant differences in exposure were identified among the income groups (Figure 3.5). Next, the Game-Howell post-hoc tests were performed to identify significant exposure differences between each pair of income groups within the two data sets. Finally, the F-values from the two Welch's ANOVA were compared to identify whether the differences were more pronounced in the GeoAir2.0 data than in the PurpleAir data. For the gender disparity, the Welch's t-test was performed. All the statistical data analyses were completed using Pandas and Matplotlib Python packages.

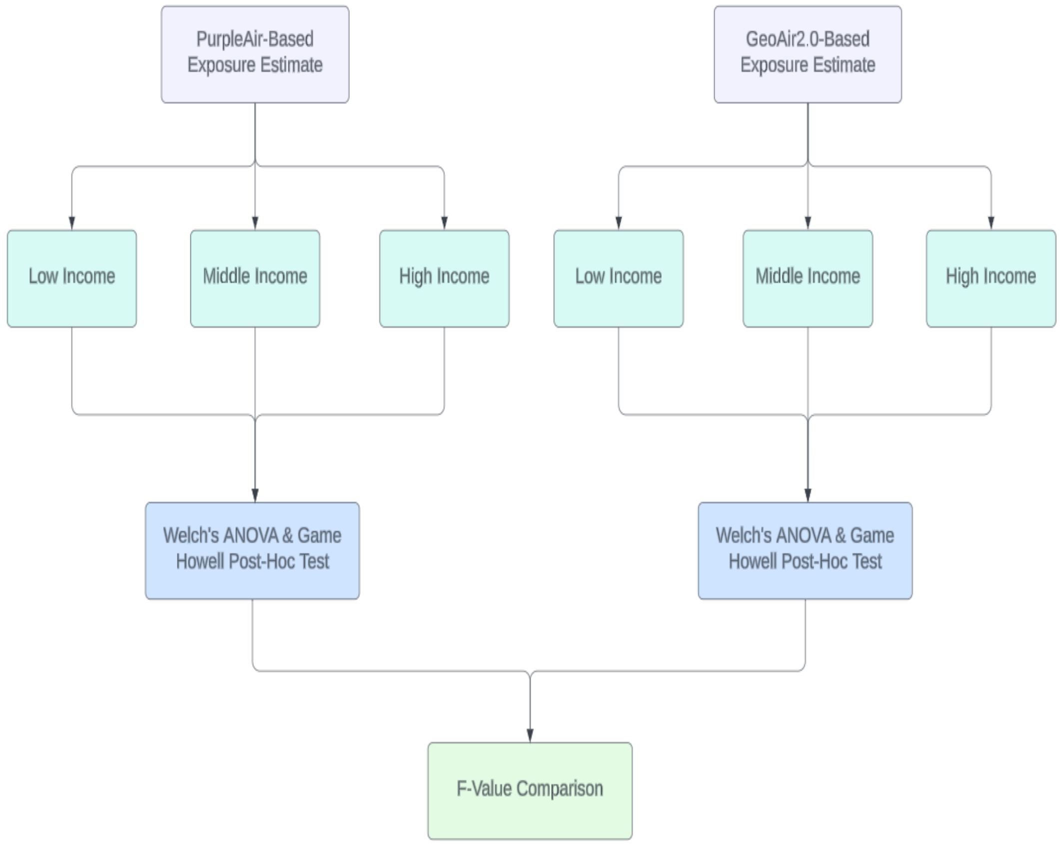

Figure 3.6 Flowchart for the statistical analyses of economic disparities in exposure: A comparison of GeoAir2.0 (mobile monitoring) and PurpleAir (stationary outdoor monitoring) data

4 CHAPTER FOUR: RESULTS

4.1 Demographic Characteristics of Participants and Average Exposure Levels

Table 4.1 shows the key demographic characteristics of the 100 research participants. Female respondents dominated the study (65%), with a significant difference between the two average exposure estimates (43.78 vs. 15.65μg/m^3). Forty-one percent of the sample were between 40-59 years, with the highest average estimate according to the PurpleAir data (51.85 μg/m^3), while their GeoAir2.0-based estimate was relatively low (11.60μg/m^3). Forty-five percent were Latino/Hispanic, experiencing an average exposure level of 60.24μg/m^3, according to PurpleAir data. This figure significantly diverges from the 16.30 μg/m^3 estimated through the GeoAir data. The overrepresentation of the Latino/Hispanic group in the sample is attributed to the initial recruitment for and the original purpose of the pilot project, upon which this current study is based, which specifically focused on this racial/ethnic group. Thirty-seven percent of the samples were from low-income households with an average exposure estimate of 45.54 (PurpleAir) vs. 21.19 (GeoAir2.0) μg/m^3.

Table 4.1 Key demographic characteristics of the study participants

Variables		Total N	Average Exposure (μg/m^3)	
		100	GeoAir2.0	PurpleAir
Gender				
	Male	35	8.55	46.26
	Female	65	15.65	43.78
Age				
	18 - 24	20	17.63	38.82
	25 - 39	34	13.95	41.52
	40 - 59	41	11.60	51.85
	60 Plus	5	2.80	30.12
Ethnicity/Race				
	Latino/Hispanic	45	16.30	60.24
	Non-Hispanic white	24	10.06	35.31
	Black	18	15.26	17.71

	Asian	13	5.14	45.47
Household Income				
	≤ $30,000	37	21.19	45.54
	$30,001–$74,999	25	11.21	32.41
	≥ $75,000	24	5.31	36.01
	PNR*	14	-	-

* PNR: Prefer not to respond.

4.2 Bias in Personal Exposure Estimates

The paired-sample t-test result revealed that there was a statistically significant difference between the GeoAir2.0- and PurpleAir-based estimates for total exposure ($t = 5.94$; $p < 0.001$). Specifically, estimates based only on outdoor data (PurpleAir) were considerably higher than the GeoAir-based estimates and consistently overestimated levels of personal exposure (Table 4.2). This indicates that even in areas and time periods with high levels of outdoor air pollution, individuals actually experienced low levels of exposure by staying indoors on average.

Table 4.2 Paired-sample t-test results

	Exposure Type	Data Source	Mean	St.D	t-value	p-value
	Total exposure	PurpleAir data	44.64	47.99	5.94	< 0.001
		GeoAir2.0 data	13.16	20.34		
	Home exposure	PurpleAir data	44.83	46.56	5.74	< 0.001
		GeoAir2.0 data	13.96	23.42		
Place	In-transit exposure	PurpleAir data	44.43	53.44	4.25	< 0.001
		GeoAir2.0 data	10.38	56.07		
	Work exposure	PurpleAir data	38.39	46.13	4.68	< 0.001
		GeoAir2.0 data	7.27	12.17		
	Other place exposure	PurpleAir data	42.32	55.70	3.01	0.003
		GeoAir2.0 data	18.43	54.85		
	Weekdays exposure	PurpleAir data	43.92	47.36	6.03	< 0.001
		GeoAir2.0 data	12.08	19.59		

Days of the week	Weekend exposure	PurpleAir data	43.99	50.38	4.75	< 0.001
		GeoAir2.0 data	16.12	29.52		

The discrepancy between the two exposure estimates was further illustrated by examining each pair of estimates (Figure 4.1). It found that the overall high-low exposure patterns across participants in the PurpleAir data did not align with those in the GeoAir data. That is, the estimates from the two datasets did not follow similar patterns. For example, a participant with an identification (ID) number of P53 seemed to experience the highest $PM_{2.5}$ exposure compared to all other participants, according to the PurpleAir data (Figure 4.1). This person encountered 240.45 μg/m^3 on average over a period of three days, which is considered “Very Unhealthy” according to the United States Environmental Protection Agency (US EPA) guidelines. However, the GeoAir2.0-based estimate revealed that the person actually experienced among the lowest exposure (4.05 μg/m^3). This sharp contrast suggests that PurpleAir-based estimates failed to account for the protective effect of being indoors when outdoor air quality was markedly poor. On the contrary, a participant with an ID number of P42 seemingly had among the low exposure (7.12 μg/m^3) when assessed using the PurpleAir data, whereas the person actually experienced an “Unhealthy” level of $PM_{2.5}$ (105.01 μg/m^3) according to the GeoAir2.0 data. This may be because the person performed indoor activities that generated a significant amount of $PM_{2.5}$ while outdoor air quality was reasonably good.

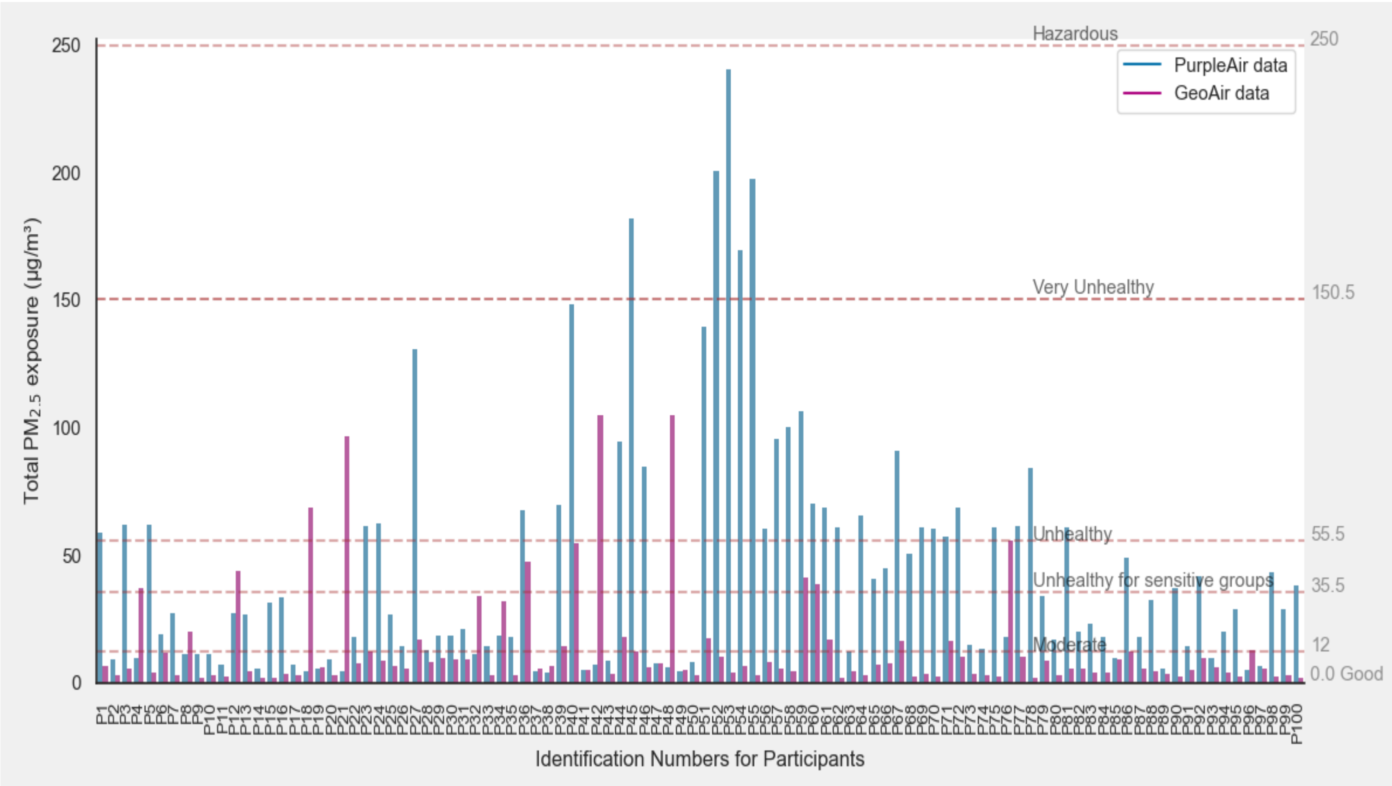

Figure 4.1 Comparing PurpleAir- and GeoAir2.0-based personal exposure estimates

This study also identified the statistically significant discrepancies between the two exposure estimates in all different types of places and time periods (Table 4.2). The PurpleAir estimates were significantly higher across all the activity places. It indicates that using the PurpleAir-based outdoor air quality data can overestimate people's actual exposure at all activity locations due to neglecting the protective effect of being indoors. The largest difference existed in-home exposure ($t = 5.74$; $p < 0.001$), while locations classified as "other places" reported the smallest t-value ($t = 3.01$; $p = 0.003$). The most pronounced contrast in-home exposure may be because GeoAir2.0 data might have almost exclusively included inside-home air quality when the location was classified as "home," while PurpleAir data included exclusively outside-home air quality at the same location. On the other hand, for locations classified as "other places," GeoAir2.0 data encompasses a mix of indoor and outdoor air quality measurements in those places due to the diverse nature of these environments, where some may be predominantly outdoors (e.g., parks and playgrounds) and others primarily indoors (e.g., restaurants and grocery stores), reducing the difference between the two data. The results also revealed that in the PurpleAir-based estimates, the highest average exposure was found at home locations (44.83 μg/m^3); in contrast, in GeoAir-based estimates, the highest average exposure was found in locations classified as "other places" (18.43 μg/m^3), probably because of the pollution-generating activities some participants engaged in those locations (Figure 4.2). Using the same data, Park et al. (2023) reported that some participants were exposed to dangerous levels of $PM_{2.5}$ during religious activities due to incense sticks burning in churches or recreational activities, such as bonfires with outdoor grilling. This study also found that workplace locations had the lowest averages for both GeoAir and PurpleAir estimates (7.27 and 38.39 μg/m^3, respectively) among all place types.

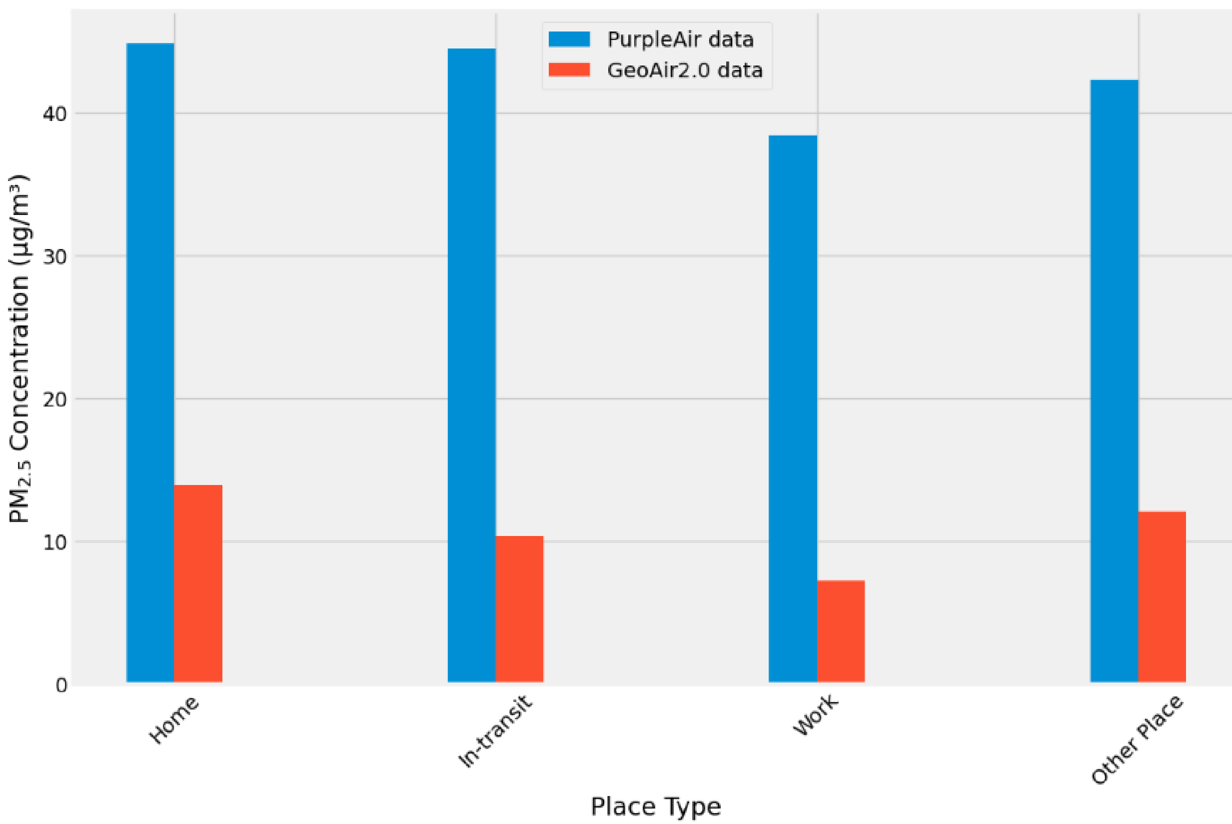

Figure 4.2 Differences between PurpleAir- and GeoAir2.0-based exposure estimates in each activity place

The difference between weekday and weekend exposure was more prominent in the GeoAir data-based estimates (12.08 vs. 16.12 μg/m^3) than in the PurpleAir data-based estimates (43.92 vs. 43.99 μg/m^3) (Figure 4.3). It may be because, during weekends, many people might have spent more time at home and performed pollution-generating indoor activities, such as cooking or vacuuming, or undertaken recreational, social, and religious activities that might have generated a significant amount of air pollution. In contrast, the difference between weekdays and weekends was not apparent in the PurpleAir-based estimates, indicating that using only outdoor data does not capture the true exposure encountered in their immediate surroundings and the impact of indoor activities frequently performed during weekends on variations in levels of exposure.

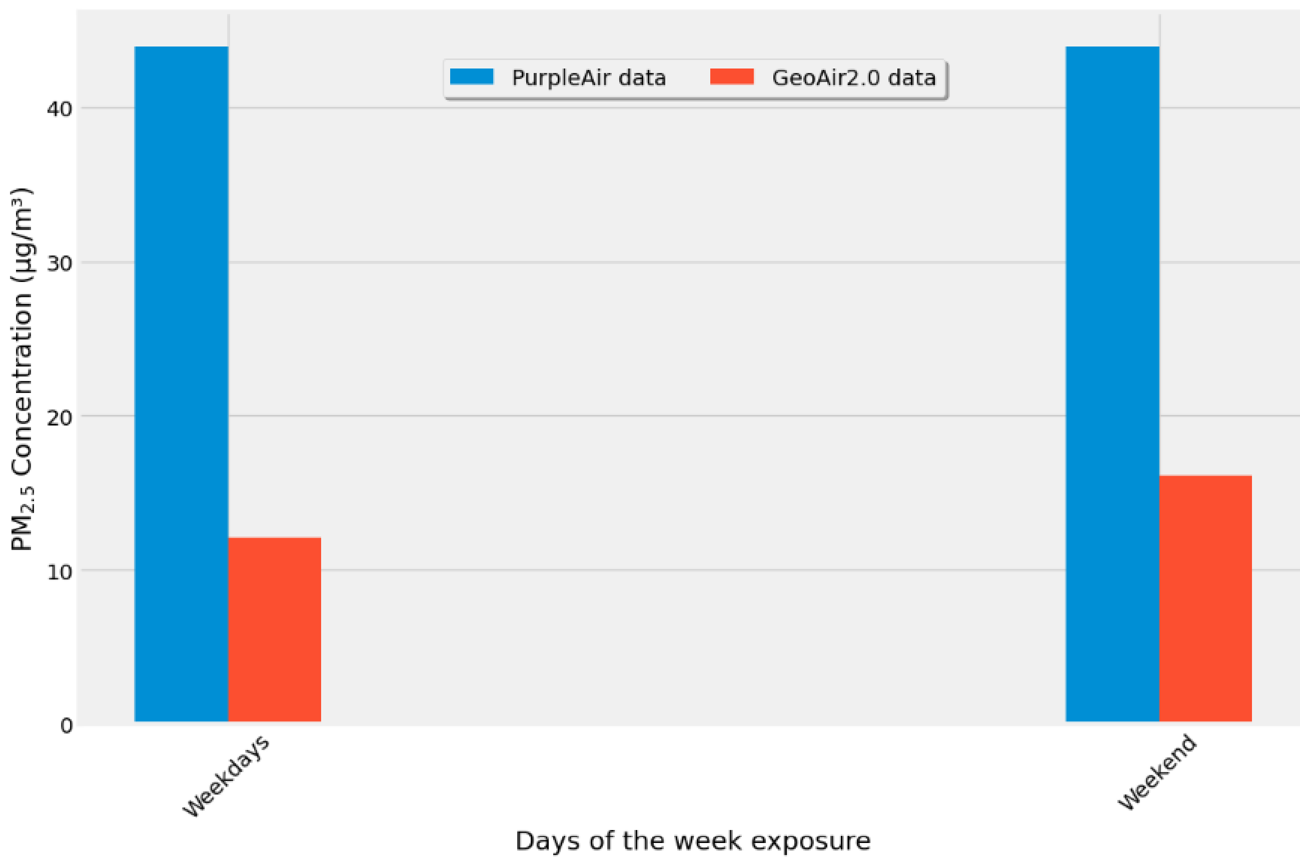

Figure 4.3 Differences between PurpleAir- and GeoAir2.0-based exposure estimates on weekdays and weekends

4.3 Bias in Assessments of Gender/Economic Disparities in Exposure

This study compared the two exposure estimates among gender groups (male and female); the finding highlighted that females had higher average exposure to air quality (15.65 $\mu g/m^3$), almost double the male exposure to air pollution in GeoAir-based estimates. In comparison, males had higher exposure in PurpleAir-based estimates (Table 4.1). The difference between the two genders was "marginally" significant only in GeoAir ($t = 4.654$; $p = 0.055$) but not in PurpleAir ($t = 3.793$; $p = 0.821$), possibly because many female participants relatively tended to spend more time at home than male participants and/or because female participants might have engaged more in air pollution-generating activities in indoor homes than male.

This study also compared the two exposure estimates for different income groups to determine whether PurpleAir- and GeoAir-based estimates led to the same conclusion about economic disparities in exposure. The Welch's ANOVA results suggested that there was no significant difference among the three income groups when using the PurpleAir data ($F = 0.756$; $p = 0.474$) (Table 4.3). In the PurpleAir-based estimates, the low-income group had the highest average (45.54 $\mu g/m^3$), followed by the high-income group (36.01 $\mu g/m^3$), and middle-income had the lowest average (32.42 $\mu g/m^3$). Conversely, the significant differences among the income groups were identified when the GeoAir data was used ($F = 7.512$; $p = 0.002$). The higher F-value for the GeoAir data indicates that the differences between the income groups are more pronounced in the GeoAir-based estimates than in those from the PurpleAir data. In the GeoAir-based estimates, the low-income group had the highest average (21.19 $\mu g/m^3$), followed by the middle-income group (11.22 $\mu g/m^3$), and the high-income group had the lowest average (5.30 $\mu g/m^3$). The distributions of the two estimates for each income group are presented in Figure 4.4.

Table 4.3 Welch's ANOVA results for income groups

Data source	Income group	Mean	St.D	F-value	p-value
PurpleAir data	Low Income	45.54	52.50	0.756	0.474
	Middle Income	32.42	31.24		
	High Income	36.01	25.19		
GeoAir2.0 data	Low Income	21.19	29.76	7.512	0.002
	Middle Income	11.22	11.55		
	High Income	5.30	4.02		

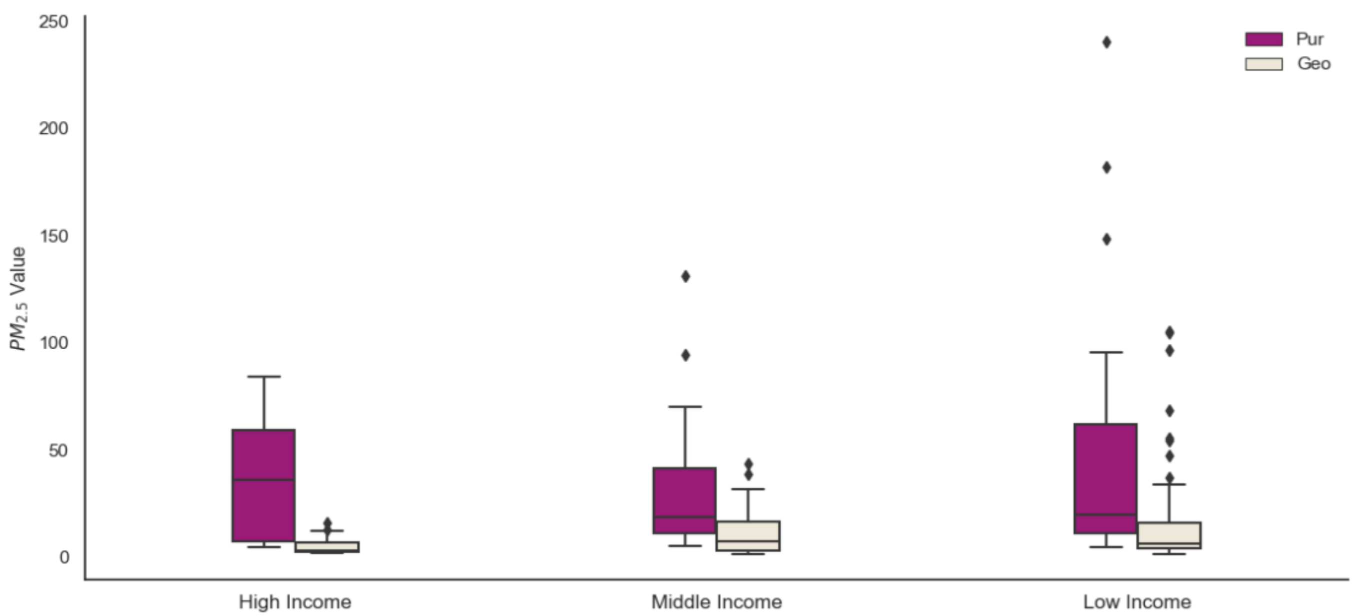

Figure 4.4 Comparing the distributions of PurpleAir- and GeoAir2.0-based $PM_{2.5}$ exposure estimates by income

The Game-Howell post hoc test was performed to identify which pairs of groups had significant differences between the mean values. The results show that the significance in Welch's ANOVA emanated from the high-low pair (p = 0.008). The high-middle pair was marginally significant, with a p-value of 0.056, and the low-middle pair was insignificant (Table 4.4). On the other hand, no pairs of groups were significant in the PurpleAir-based estimates, and all these pairs had very high p-values. These results suggest that using only outdoor air quality data in exposure assessments may not accurately reveal the magnitude of disparities in exposure that may exist among different socioeconomic groups. Incorporating indoor air quality into exposure assessments can reveal a hidden portion of disparities by capturing the unequal distribution of indoor sources and differences in indoor environmental quality, which is neglected in outdoor-focused assessments.

Table 4.4 Game-Howell posthoc test results for income groups

Data source	**Pair of groups**	**p-value**
PurpleAir data	High - Low Income	0.612
	High - Middle Income	0.897
	Low - Middle Income	0.439
GeoAir2.0 data	High - Low Income	0.008
	High - Middle Income	0.056
	Low - Middle Income	0.166

5 CHAPTER FIVE: DISCUSSION OF FINDINGS, RESEARCH LIMITATIONS, AND CONCLUSIONS

5.1 Discussion

This study revealed the vast discrepancy between personal exposure estimates derived from mobile sensor data and stationary outdoor monitoring data (i.e., GeoAir2.0 data vs. PurpleAir data). The PurpleAir-based assessments tended to overestimate individuals' actual daily exposure levels when outdoor air quality was relatively poor and underestimate them when outdoor air quality was good. For some participants, exposure estimates derived from GeoAir2.0 data were among the lowest even though their exposure in PurpleAir data was among the highest, and vice versa. These findings suggest that exposure assessments based exclusively on outdoor air quality data may lead to biased exposure estimates, underscoring the necessity of incorporating indoor air quality into these assessments by collecting indoor and outdoor air quality data using mobile sensors. Moreover, the analyses showcasing variations in exposure across weekends versus weekdays and among different types of places further highlight the critical role of indoor environments in determining personal exposure. These outcomes align with the Ma et al. (2020) study, which also identified the significant discrepancies between estimates derived from mobile sensor data and those from national monitoring data in Beijing, China.

Additionally, this study demonstrates that neglecting indoor air quality may underrepresent the extent of social disparities in exposure. Notably, it found considerable variations in exposure levels among different income groups when employing the GeoAir2.0 data, with lower-income groups experiencing the highest average exposure—a disparity not captured by PurpleAir data. The study also highlights that neglecting indoor air quality does not reveal the true extent of gender disparities in exposure — GeoAir data-based exposure estimates revealed gender disparities in exposure, while

PurpleAir data-based exposure estimates did not. These results indicate that using low-cost portable air sensors to obtain high-resolution, personalized air quality data collected in both indoor and outdoor microenvironments can significantly reduce bias in personal exposure estimates and uncover the often-overlooked disparities indoors. This conclusion is consistent with the conclusions drawn in several previous studies (Beko et al., 2015; Ma et al., 2020; Park et al., 2023).

5.2 Limitations and Future Research Directions

This study has some noteworthy limitations. First, this research primarily focused on small to mid-size cities (e.g., Greenville) and some neighboring rural areas across ENC and the Raleigh/Durham area, using the data collected during the cold season, a time when outdoor air pollution tends to be higher compared to warmer seasons. Consequently, the observed overall trend of personal exposure being overestimated when relying solely on outdoor air quality data may not hold true across different times of the year or across the full spectrum of the most remote rural areas to the heavily industrialized urban centers. Further studies are required to explore the extent and characteristics of bias in exposure assessments across different times and regions.

Second, this study assumes that each participant's 3-day data represents their typical weekday and weekend exposure patterns. However, the data might not have accounted for other sources of $PM_{2.5}$ that they occasionally encounter and long-term trends in exposure. Future research may consider collecting data for an extended period to capture these occasional exposures and seasonal variations in biases. Third, this study specifically targeted bias in assessments of economic and gender disparities because the samples were more evenly distributed across different groups, ensuring a sufficient number of participants in each category. Statistical analysis on racial disparities was not conducted due to concerns about the reliability of the results. The Latino/Hispanic group was oversampled due to the study design of the original pilot project while there was an inadequate number of participants from the

Black population to allow for meaningful comparisons. Future research may employ a more effective sampling strategy to obtain a representative sample with an adequate number of participants from each racial/ethnic group to facilitate the identification of bias in assessments of racial disparities in exposure.

Lastly, the discrepancy in the two exposure estimates (GeoAir2.0 and PurpleAir) might have partially stemmed from the differences in accuracy between the two low-cost sensors as well as the measurement errors. In a related context, the GeoAir-based exposure estimates were based on direct measurements from sensors, while PurpleAir-based estimates were derived from the geostatistical model outputs. Although this study included data points in the surrounding areas to improve accuracy, estimations at locations where the available data points were insufficient might have had higher errors. Future research might consider using modern machine-learning models, such as Gaussian Process regression, which might provide more accurate estimation results. Nonetheless, the magnitude of the divergence identified in this study suggests that such factors alone are unlikely to account for the majority of the observed large discrepancy between the two data.

5.3 Conclusions

This study contributes to the literature by identifying potential inferential errors in exposure or health risk assessments or environmental justice studies that exclude indoor air pollution. Although this study found that assessments based solely on outdoor air quality data resulted in overestimated personal exposure, this conclusion does not undermine the necessity of air pollution management in both indoor and outdoor environments because GeoAir data revealed that people were exposed to extremely high levels of air pollution for a short term (several minutes or hours per day), while the calculated average values concealed these short-term variations. Rather, this study suggests that the bias is significant and non-negligible, emphasizing the importance of integrating air quality data collected from both indoor and outdoor environments to mitigate bias in exposure estimates despite the expensive and time-consuming process of acquiring real-time air sampling data from individuals. It calls for developing innovative data collection strategies to collect personalized, high-resolution air pollution data across a large geographic area and research methods that integrate such data. A more accurate exposure assessment will contribute to enhancing an understanding of health risks associated with air pollution and developing more nuanced intervention strategies and policies for mitigating the risks and environmental health disparities.

BIBLIOGRAPHY

Abdel-Salam, M. M. M. (2022). Relationship between residential indoor air quality and socioeconomic factors in two urban areas in Alexandria, Egypt. *Building and Environment*, *207*, 108425.

Adamkiewicz, G., Zota, A. R., Fabian, M. P., Chahine, T., Julien, R., Spengler, J. D., & Levy, J. I. (2011). Moving Environmental Justice Indoors: Understanding Structural Influences on Residential Exposure Patterns in Low-Income Communities. *American Journal of Public Health*, *101*(S1), S238–S245.

Ambient (outdoor) Air Pollution. (2022, December 19). Retrieved February 16, 2023, from https://www.who.int/news-room/fact-sheets/detail/ambient-(outdoor)-air-quality-and-health

Bales, E., Nikzad, N., Quick, N., Ziftci, C., Patrick, K., & Griswold, W. G. (2019). Personal pollution monitoring: mobile real-time air quality in daily life. *Personal and Ubiquitous Computing*, *23*(2), 309–328.

Barkjohn, K. K., Holder, A. L., Frederick, S. G., & Clements, A. L. (2022). Correction and accuracy of PurpleAir $PM_{2.5}$ measurements for extreme wildfire smoke. *Sensors*, *22*(24), 9669.

Breen, M. S., Long, T., Schultz, B. D., Rt, W., Richmond-Bryant, J., Breen, M., Langstaff, J., Devlin, R. B., Schneider, A., Burke, J., Batterman, S., & Meng, Q. Y. (2015). Air Pollution Exposure Model for Individuals (EMI) in Health Studies: Evaluation for ambient PM2.5 in Central North Carolina. *Environmental Science & Technology*, *49*(24), 14184–14194.

Brown, T. R., Dassonville, C., Derbez, M., Ramalho, O., Kirchner, S., Crump, D., & Mandin, C. (2015b). Relationships between socioeconomic and lifestyle factors and indoor air quality in French dwellings. *Environmental Research*, *140*, 385–396. Buonanno, G., Stabile, L., & Morawska, L. (2014). Personal exposure to ultrafine particles: The influence of time-activity patterns. *Science of the Total Environment*, *468–469*, 903–907.

Caplin, A., Ghandehari, M., Lim, C., Glimcher, P. W., & Thurston, G. D. (2019). Advancing environmental exposure assessment science to benefit society. *Nature Communications*, *10*(1).

Chen, P., & Lin, Y. (2022). Exposure assessment of $PM_{2.5}$ using smart spatial interpolation on regulatory air quality stations with clustering of densely-deployed microsensors. *Environmental Pollution, 292*, 118401.

Chen, B., Song, Y., Jiang, T., Chen, Z., Huang, B., & Xu, B. (2018b). Real-time estimation of population exposure to $PM_{2.5}$ using mobile- and Station-Based big data. *International Journal of Environmental Research and Public Health/International Journal of Environmental Research and Public Health*, *15*(4), 573.

Dias, D., & Tchepel, O. (2018). Spatial and temporal dynamics in air pollution Exposure assessment. *International Journal of Environmental Research and Public Health (Online)*, *15*(3), 558.

Doris, M., Daley, C., Zalzal, J., Chesnaux, R., Minet, L., Kang, M., Caron-Beaudoin, É., MacLean, H. L., & Hatzopoulou, M. (2024). Modelling spatial & temporal variability of air pollution in an area of unconventional natural gas operations. *Environmental Pollution (1987)*, 123773.

Dwivedi, S., Zehra, F., Masih, J., Gupta, T., & Lawrence, A. J. (2024). Investigating the temporal dynamics of sub-micron particles and particle-bound transition metals in the indoor air of a metropolitan city. *Environmental Geochemistry and Health*, *46*(2).

Jiang, Y., Zhu, X., Chen, C., Ge, Y., Wang, W., Zhao, Z., Cai, J., & Kan, H. (2021). On-field test and data calibration of a low-cost sensor for fine particles exposure assessment. *Ecotoxicology and Environmental Safety*, *211*, 111958.

Jing, Z., Gladson, L., & Cromar, K. (2018). A novel environmental justice Indicator for managing local air pollution. *International Journal of Environmental Research and Public Health*, *15*(6), 1260.

Kampa, M., & Castanas, E. (2008). Human health effects of air pollution. *Environmental Pollution*, *151*(2), 362–367.

Kerckhoffs, J., Hoek, G., Messier, K. P., Brunekreef, B., Meliefste, K., Klompmaker, J. O., & Vermeulen, R. (2016b). Comparison of Ultrafine Particle and Black Carbon Concentration Predictions from a Mobile and Short-Term Stationary Land-Use Regression Model. *Environmental Science & Technology*, *50*(23), 12894–12902.

Kim, J., & Kwan, M. (2021). Assessment of sociodemographic disparities in environmental exposure might be erroneous due to neighborhood effect averaging: Implications for environmental inequality research. *Environmental Research*, *195*, 110519.

Klepeis, N. E., Bellettiere, J., Hughes, S., Nguyen, B., Berardi, V., Liles, S., Obayashi, S., Hofstetter, C. R., Blumberg, E. J., & Hovell, M. F. (2017). Fine particles in homes of predominantly low-income families with children and smokers: Key physical and behavioral determinants to inform indoor-air-quality interventions. *PLOS ONE*, *12*(5), e0177718.

Kwan, M. P. (2012a). The Uncertain Geographic Context Problem. *Annals of the Association of American Geographers*, *102*(5), 958–968.

Kwan, M. P. (2013a). Beyond Space (As We Knew It): Toward Temporally Integrated Geographies of Segregation, Health, and Accessibility. *Annals of the Association of American Geographers, 103*(5), 1078–1086.

Ma, J., Tao, Y., Kwan, M. P., & Chai, Y. (2020b). Assessing Mobility-Based Real-Time Air Pollution Exposure in Space and Time Using Smart Sensors and GPS Trajectories in Beijing. *Annals of the American Association of Geographers*, *110*(2), 434–448.

Masiol, M., Zíková, N., Chalupa, D. C., Rich, D. Q., Ferro, A. R., & Hopke, P. K. (2018). Hourly land-use regression models based on low-cost PM monitor data. *Environmental Research*, *167*, 7–14.

Minet, L., Gehr, R., & Hatzopoulou, M. (2017). Capturing the sensitivity of land-use regression models to short-term mobile monitoring campaigns using air pollution micro-sensors. *Environmental Pollution, 230,* 280–290.

North Carolina Air Quality Index (AQI) and USA Air Pollution. (2O23) *IQAir*, Retrieved April 16, 2023, from https://www.iqair.com/us/usa/north-carolina.

Liu, J., Banerjee, S., Oroumiyeh, F., Shen, J., Del Rosario, I., Lipsitt, J., Paulson, S., Ritz, B., Su, J., Weichenthal, S., Lakey, P., Shiraiwa, M., Zhu, Y., & Jerrett, M. (2022). Co-kriging with a low-cost sensor network to estimate spatial variation of brake and tire-wear metals and oxidative stress potential in Southern California. *Environment International*, *168*, 107481.

Lu, Y. (2021). Beyond air pollution at home: Assessment of personal exposure to $PM_{2.5}$ using activity-based travel demand model and low-cost air sensor network data. *Environmental Research*, p. *201*, 111549.

Ma, J., Tao, Y., Kwan, M. P., & Chai, Y. (2020b). Assessing Mobility-Based Real-Time Air Pollution Exposure in Space and Time Using Smart Sensors and GPS Trajectories in Beijing. *Annals of the American Association of Geographers*, *110*(2), 434–448.

Mocumbi, A. O., Stewart, S., Patel, S., & Al-Delaimy, W. K. (2019). Cardiovascular Effects of Indoor Air Pollution from Solid Fuel: Relevance to Sub-Saharan Africa. *Current Environmental Health Reports*, *6*(3), 116–126.

Nyhan, M., Grauwin, S., Britter, R., Misstear, B., McNabola, A., Laden, F., Barrett, S. R. H., & Ratti, C. (2016). “Exposure Track”—The Impact of Mobile-Device-Based mobility Patterns on quantifying population exposure to air pollution. *Environmental Science & Technology*, *50*(17), 9671–9681.

Omelekhina, Y., Eriksson, A., Canonaco, F., Prévôt, A. S. H., Nilsson, P., Isaxon, C., Pagels, J., & Wierzbicka, A. (2020). Cooking and electronic cigarettes leading to large differences between indoor and outdoor particle composition and concentration measured by aerosol mass spectrometry. *Environmental Science. Processes & Impacts*, *22*(6), 1382–1396.

Opejin, A. (2024a). Step-By-Step Procedures to Download PurpleAir Sensors Data from PurpleAir Network Using PurpleAir API and Data Download Tools. Medium. [https://medium.com/@opejinabdulahiay/step-by-step-on-how-to-download-purpleair-sensors-data-from-purpleair-network-e6ed6ccb27b8]

Opejin, A. (2024b). Air-Pollution Estimation Model. https://github.com/OpejinAbdulahi/Air-Pollution-Estimation-Model

Park, Y. M. (2022). A GPS-enabled portable air pollution sensor and web-mapping technologies for field-based learning in health geography. *Journal of Geography in Higher Education*, *46*(2), 241–261.

Park, Y. M. (2020). Assessing personal exposure to traffic-related air pollution using individual travel-activity diary data and an on-road source air dispersion model. *Health & Place*, *63*, 102351.

Park, Y. M., Chavez, D., Sousan, S., Figueroa-Bernal, N., Alvarez, J. R., & Rocha-Peralta, J. (2023). Personal exposure monitoring using GPS-enabled portable air pollution sensors: A strategy to promote citizen awareness and behavioral changes regarding indoor and outdoor air pollution. *Journal of Exposure Science and Environmental Epidemiology*, 33(3), 347–357.

Park, Y. M., & Kwan, M. P. (2017). Individual exposure estimates may be erroneous when spatiotemporal variability of air pollution and human mobility are ignored. *Health & Place*, *43*, 85–94.

Park, Y. M., Sousan, S., Streuber, D., & Zhao, K. (2021). GeoAir—A Novel Portable, GPS-Enabled, Low-Cost Air-Pollution Sensor: Design Strategies to Facilitate Citizen Science Research and Geospatial Assessments of Personal Exposure. *Sensors*, *21*(11), 3761.

PurpleAir - *Community sensor map*. (n.d.). Mojave Desert Air Quality Management District. Retrieved March 4, 2023, from https://www.mdaqmd.ca.gov/air-quality/purple-air-community-monitor-map and https://map.purpleair.com/1/mAQI/a10/p604800/cC0#5.29/36.792/-78.778

Scharer, J. (2001). How does the East compare to the rest of North Carolina?

Son, J., Bell, M. L., & Lee, J. (2010). Individual exposure to air pollution and lung function in Korea: Spatial analysis using multiple exposure approaches. *Environmental Research*, *110*(8), 739–749.

Son Bekö, G., Kjeldsen, B. U., Olsen, Y., Schipperijn, J., Wierzbicka, A., Karottki, D. G., Toftum, J., Loft, S., & Clausen, G. (2015). Contribution of various microenvironments to the daily personal exposure to ultrafine particles: Personal monitoring coupled with GPS tracking. *Atmospheric Environment*, *110*, 122–129.

Sousan, S., Regmi, S., & Park, Y. M. (2021b). Laboratory Evaluation of Low-Cost Optical Particle Counters for Environmental and Occupational Exposures. *Sensors*, *21*(12), 4146.

Stavroulas, I., Grivas, G., Michalopoulos, P., Liakakou, E., Bougiatioti, A., Kalkavouras, P., Fameli, K. M., Hatzianastassiou, N., Mihalopoulos, N., & Gerasopoulos, E. (2020). Field evaluation of Low-Cost PM sensors (Purple Air PA-II) under variable urban air quality conditions, in Greece. *Atmosphere*, *11*(9), 926. https://doi.org/10.3390/atmos11090926

Steinle, S., Reis, S., Sabel, C. E., Semple, S., Twigg, M. M., Braban, C. F., Leeson, S. R., Heal, M. R., Harrison, D., Lin, C., & Wu, H. (2015c). Personal exposure monitoring of $PM_{2.5}$ in indoor and outdoor microenvironments. *Science of the Total Environment*, *508,* 383–394.

Streuber, D., Park, Y. M., & Sousan, S. (2022). Laboratory and Field Evaluations of the GeoAir2 Air Quality Monitor for Use in Indoor Environments. *Aerosol and Air Quality Research*, *22*(8), 220119.

Tong, X., & Ho, K. (2018). Indoor air pollutants exposure and determinant factors controlling households air quality for elderly people in Hong Kong. *Environmental Health Perspectives. Supplements*, *2018*(1).

Vardoulakis, S., Giagloglou, E., Steinle, S., Davis, A., Sleeuwenhoek, A., Galea, K. S., Dixon, K., & Crawford, J. (2020). Indoor exposure to selected air pollutants in the home environment: a systematic review. *International Journal of Environmental Research and Public Health (Online)*, *17*(23), 8972.

Vicente, E. D., Calvo, A., Sainnokhoi, T., Kováts, N., De La Campa, A. M. S., De La Rosa, J., Oduber, F., Nunes, T., Fraile, R., Tomé, M., & Alves, C. (2024). Indoor PM from residential coal combustion: Levels, chemical composition, and toxicity. *Science of the Total Environment*, 170598.

Vu, T. V., Stewart, G., Kitwiroon, N., Lim, S., Barratt, B., Kelly, F. J., Thompson, R., Smith, R. B., Toledano, M. B., & Beevers, S. (2022). Assessing the contributions of outdoor and indoor sources to air quality in London homes of the SCAMP cohort. *Building and Environment*, p. *222*, 109359.

Wang, J., Du, W., Lei, Y., Chen, Y., Li, X., Mao, K., Tao, S., & Pan, B. (2023). Quantifying the dynamic characteristics of indoor air pollution using real-time sensors: Current status and future implication. *Environment International*, *175*, 107934.

Yoo, E., Pu, Q., Eum, Y., & Jiang, X. (2021). The impact of individual mobility on Long-Term Exposure to Ambient $PM_{2.5}$: Assessing effect modification by travel patterns and spatial variability of $PM_{2.5}$. *International Journal of Environmental Research and Public Health (Online)*, *18*(4), 2194.

Yu, M., Zhang, S., Ning, H., Li, Z., & Zhang, K. (2024). Assessing the 2023 Canadian wildfire smoke impact in Northeastern US: Air quality, exposure and environmental justice. *Science of the Total Environment*, p. *926*, 171853.

Yu, M., Zhang, S., Zhang, K., Yin, J., Varela, M., & Miao, J. (2023). Developing high-resolution $PM_{2.5}$ exposure models by integrating low-cost sensors, automated machine learning, and big human mobility data. *Frontiers in Environmental Science*, *11*.

APPENDIX A. IRB Approval

EAST CAROLINA UNIVERSITY
University & Medical Center Institutional Review Board
4N-64 Brody Medical Sciences Building· Mail Stop 682
600 Moye Boulevard · Greenville, NC 27834
Office **252-744-2914** · Fax **252-744-2284** ·
rede.ecu.edu/umcirb/

Notification of Initial Approval: Expedited

From: Biomedical IRB
To: Sinan Sousan
CC:
Date: 7/26/2023
Re: UMCIRB 23-001348
Assessing personal exposure in various indoor and outdoor environments using low-cost air pollution sensors

I am pleased to inform you that your Expedited Application was approved. Approval of the study and any consent form(s) occurred on 7/25/2023. The research study is eligible for review under expedited category # 5. The Chairperson (or designee) deemed this study no more than minimal risk.

As the Principal Investigator you are explicitly responsible for the conduct of all aspects of this study and must adhere to all reporting requirements for the study. Your responsibilities include but are not limited to:

1. Ensuring changes to the approved research (including the UMCIRB approved consent document) are initiated only after UMCIRB review and approval except when necessary to eliminate an apparent immediate hazard to the participant. All changes (e.g. a change in procedure, number of participants, personnel, study locations, new recruitment materials, study instruments, etc.) must be prospectively reviewed and approved by the UMCIRB before they are implemented;

2. Where informed consent has not been waived by the UMCIRB, ensuring that only valid versions of the UMCIRB approved, date-stamped informed consent document(s) are used for obtaining informed consent (consent documents with the IRB approval date stamp are found under the Documents tab in the ePIRATE study workspace);

3. Promptly reporting to the UMCIRB all unanticipated problems involving risks to participants and others;

4. Submission of a final report application to the UMICRB prior to the expected end date provided in the IRB application in order to document human research activity has ended and to provide a timepoint in which to base document retention; and

5. Submission of an amendment to extend the expected end date if the study is not expected to be completed by that date. The amendment should be submitted 30 days prior to the UMCIRB approved expected end date or as

soon as the Investigator is aware that the study will not be completed by that date.

The approval includes the following items:

Name	Description
Activity_Travel_Diary.docx	Interview/Focus Group Scripts/Questions
Activity_Travel_Diary_instruction.docx	Interview/Focus Group Scripts/Questions
Air Pollution Monitoring Device Instructions.docx	Interview/Focus Group Scripts/Questions
Awareness and Behavior Survey	Surveys and Questionnaires
Awareness_Survey_Spanish	Surveys and Questionnaires
Consent form for Data Collection #2	Consent Forms
Demographic_English_0908.docx	Surveys and Questionnaires
Demographic_Spanish	Surveys and Questionnaires
Follow-up interview questions_English	Surveys and Questionnaires
Follow-up interview questions_Spanish	Surveys and Questionnaires
Informed Consent Form_revised for Amendment	Consent Forms
Informed Consent Form_revised for Amendment (Spanish)	Consent Forms
Participant Information.docx	Interview/Focus Group Scripts/Questions
research protocol	Study Protocol or Grant Application
Travel Diary revised for Amendment (English)	Surveys and Questionnaires
Travel Diary revised for Amendment (Spanish)	Surveys and Questionnaires

For research studies where a waiver or alteration of HIPAA Authorization has been approved, the IRB states that each of the waiver criteria in 45 CFR 164.512(i)(1)(i)(A) and (2)(i) through (v) have been met. Additionally, the elements of PHI to be collected as described in items 1 and 2 of the Application for Waiver of Authorization have been determined to be the minimal necessary for the specified research.

The Chairperson (or designee) does not have a potential for conflict of interest on this study.