COMMENTARY

HEI Improved Exposure Assessment Studies Review Panel

HEI

Research Report 228, *Optimizing Air Pollution Exposure Assessment with Application to Cognitive Function*, by L. Sheppard et al.

INTRODUCTION

Outdoor air pollution is a major global public health concern. There is now broad expert consensus that exposure to ambient air pollution causes an array of adverse health effects, based on evidence from a large body of scientific literature that has grown exponentially since the mid-1990s (IARC 2016; US EPA 2016, 2019, 2022; WHO 2021).

The assessment of long-term exposure to ambient air pollution for epidemiological studies, however, remains challenging. Early cohort studies characterized exposure by assigning the average concentration measured at one or a few central sites within a city to each participant from the city (Dockery et al. 1993; Pope et al. 2002). Fixed-site networks — even those in North America and Western Europe — continue to have relatively limited spatial coverage in many areas, particularly in suburban and rural locations, and insufficient density to capture small-scale (within-city) variations of air pollution (Roque et al. 2025).

Recent developments in measurement technologies and modeling approaches have increasingly been used to estimate long-term air pollution exposure at finer spatial scales for epidemiological studies of large populations. Advances include novel air pollution sensors, mobile monitoring, satellite data, hybrid models, and machine-learning approaches (Hoek 2017). Even with those advances, important limitations and challenges remain when assessing long-term air pollution exposure, particularly for pollutants that vary widely across space and time.

In 2019, HEI issued Request for Applications 19-1, *Applying Novel Approaches to Improve Long-Term Exposure Assessment of Outdoor Air Pollution for Health Studies* (see Preface). Its goal was to develop and apply scalable novel approaches to improve assessments of long-term exposures to outdoor air pollutants that vary widely in space and time — such as ultrafine particles (UFPs*), black carbon (BC), and nitrogen dioxide (NO_2). Studies were intended to evaluate exposure measurement error quantitatively and determine how exposure assessment approaches might influence the health estimates.

Dr. Sheppard and colleagues proposed to advance the understanding of exposure assessment study design features, including a comparison of health estimates derived from those features. The HEI Research Committee recommended the study for funding because of the systematic evaluation of sampling designs to provide guidance to other researchers. They also appreciated the inclusion of UFPs and the application of exposure estimates to cognitive function, as it is an emerging health outcome.

This Commentary provides the HEI Improved Exposure Assessment Studies Review Panel's evaluation of the study. It is intended to aid the sponsors of HEI and the public by highlighting the study's strengths and limitations, and by placing the results presented in the Investigators' Report into a broader scientific and regulatory context.

SCIENTIFIC AND REGULATORY BACKGROUND

Traffic-related air pollution (TRAP) continues to be an important risk factor for poor health worldwide, with the highest exposures in urban settings and at residences near busy roadways (HEI 2022). TRAP is a complex mixture of gases and particles resulting from the use of motor vehicles. Motor vehicles emit various pollutants, including NO_2, BC, and UFPs (HEI 2022). Exposure assessment of those pollutants is challenging because they are characterized by high spatial and temporal variability.

Epidemiological studies have used different approaches to address those challenges. Researchers have increasingly used mobile monitoring in recent years by affixing monitoring instruments to vehicles and making measurements while systematically and repeatedly traveling a road network. Mobile monitoring strategies can involve mobile measurements made while driving predefined routes, or repeated short-term measurements made while in a vehicle parked at various roadside locations. Data collected through mobile monitoring have been used to develop land use regression models and other air pollution maps (e.g., Apte et al. 2024; Hatzopoulou et al. 2017; Kerckhoffs et al. 2016; Messier et al. 2018). Air pollution maps estimated from such monitoring are being increasingly applied in epidemiological studies (e.g., Alexeef et al. 2018; Downward et al. 2018).

Dr. Liane Sheppard's 3-year study, "Optimizing Exposure Assessment for Inference about Air Pollution Effects with Application to the Aging Brain," began in September 2020. Total expenditures were $800,000. The draft Investigators' Report from Sheppard and colleagues was received for review in January 2024. A revised report, received in September 2024, was accepted for publication in September 2024. During the review process, the HEI Improved Exposure Assessment Studies Review Panel and the investigators had the opportunity to exchange comments and clarify issues in the Investigators' Report and the Panel's Commentary. Review Committee member Sara D. Adar was not involved in the review of this report due to a conflict of interest.

* A list of abbreviations and other terms appears at the end of this volume.

In addition, low-cost sensors are increasingly being used in exposure assessment for health studies. They can be deployed on mobile platforms or can supplement fixed-site monitoring networks to develop exposure models, or they can enable simultaneous individual-level air pollution measurements to estimate personal exposure (e.g., Larkin and Hystad 2017, Morawska et al. 2018).

Exposure models are applied in epidemiological studies that underpin the air quality standards and guidelines. Governments in the United States and Europe have recently moved toward more stringent fine particulate matter ($PM_{2.5}$) annual standards — 9 and 10 µg/m³, respectively — which align more closely with the 2021 WHO Air Quality Guidelines of 5 µg/m³. A more stringent annual standard was also set in Europe for NO_2 (**Commentary Table 1**).

There are no specific ambient air quality standards or guidelines for UFPs and BC, and regulatory agencies do not commonly measure them. Hence, international or national standard methods to characterize them have not been established (HEI 2010; HEI Review Panel on Ultrafine Particles 2013). Although no air quality guidelines have been developed for UFPs and BC, the WHO has provided "good practice statements" for these pollutants geared toward additional monitoring, mitigation, and epidemiological research (WHO 2021).

As noted earlier, important limitations and challenges remain when predicting long-term air pollution exposure to pollutants that vary highly in space and time. The current study compared the performance of different exposure assessment study design features on long-term exposure estimates of UFPs, NO_2, and $PM_{2.5}$ in Seattle, Washington.

STUDY OBJECTIVES

The overarching aim of Dr. Sheppard's study was to advance the understanding of exposure assessment study design and analysis features for air pollution and health studies. The investigators specified the following four study aims:

1. Identify key design choices to improve long-term average exposure predictions using mobile monitoring campaigns and fixed-site networks of low-cost sensors
2. Develop annual average TRAP exposure predictions from mobile monitoring data using advanced statistical methods
3. Determine the impact of sampling designs and analytical approaches on the health estimates
4. Address the overall value of incorporating novel exposure data collection and modeling by comparing the logistical features (cost and time) of using different sampling designs and analysis choices

SUMMARY OF APPROACH AND METHODS

Dr. Sheppard and colleagues compared the performance of different exposure assessment study design features on long-term exposure estimates in Seattle, Washington. In a cohort study, the investigators evaluated how various approaches to air pollution sampling affected exposure prediction and health estimates. Most analyses focused on UFP data from a previously conducted mobile roadside monitoring campaign in 2019–2020. In that campaign, short-term measurements were made from a parked vehicle at 309 roadside locations, with about 30 visits made to each site. For the present study, the investigators also conducted analyses of UFPs using mobile on-road monitoring data for a total of 5,878 road segments. Each 100-meter segment was visited an average of 28 times. Further analyses were conducted on $PM_{2.5}$ and NO_2 concentrations collected using low-cost sensors at about 115 fixed monitoring sites in 2017–2020, combined with regulatory monitoring data from a much longer time period (**Commentary Table 2**).

The investigators used either the full dataset or subsets of measurements to develop annual average exposure estimates using a suite of models, including universal kriging, spatiotemporal models, machine learning, and other advanced statistical models (Commentary Figure 1). The study leveraged

Commentary Table 1. Annual NO_2 and $PM_{2.5}$ Standards in the US, EU, and WHO

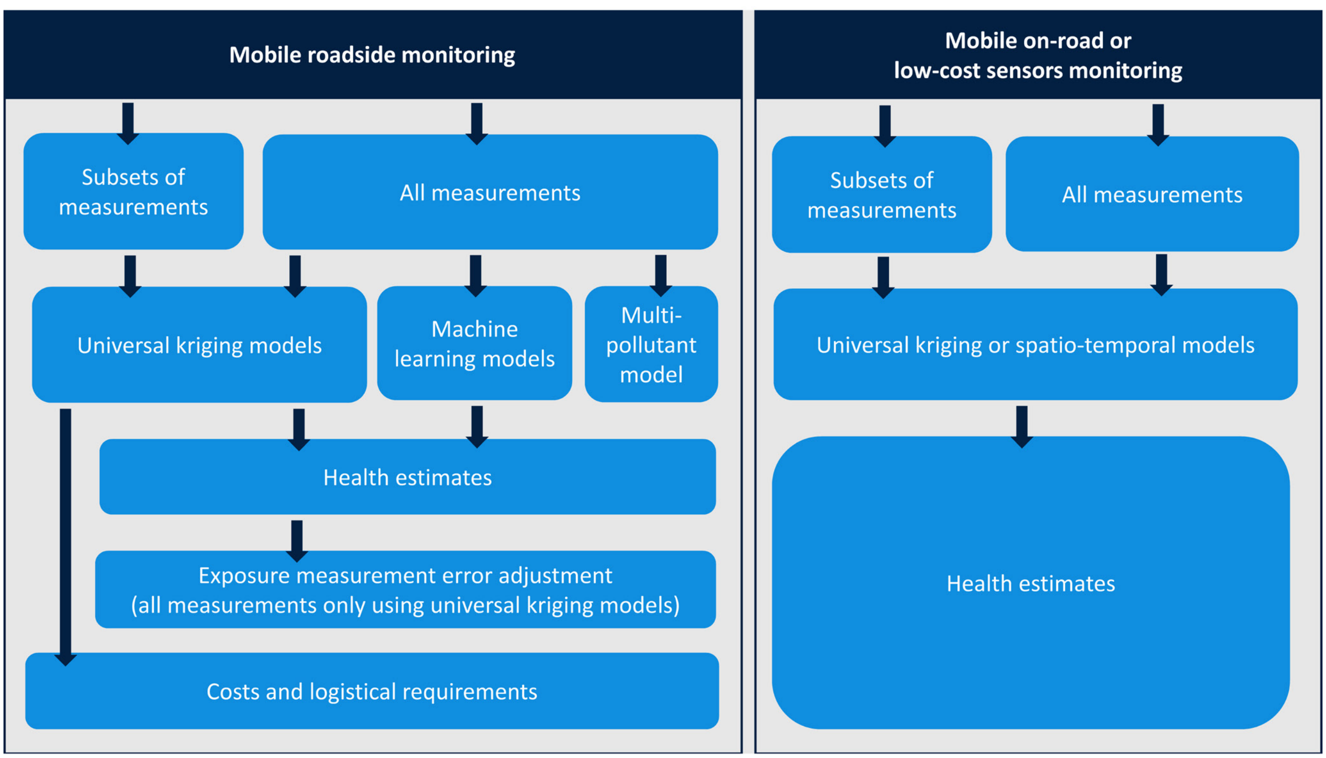
Organization	Annual $PM_{2.5}$ (µg/m³)	Annual NO_2 (µg/m³)	Notes
US EPA (2024)	9	100	NAAQS
US EPA (Previous)	12	100	Previous NAAQS
EU (2024)	10	20	Limit value for 2030
EU (Previous)	25	40	Previous limit value
WHO (2021)	5	10	Air Quality Guidelines
WHO (Previous)	10	40	Previous Air Quality Guidelines

NAAQS = National Ambient Air Quality Standards.

Commentary Table 2. Summary of Data and Analyses Conducted

Monitoring Data	Main Pollutants (UFP Device[a])	Exposure Model	Type of Analysis	Aim	Report Chapter
Mobile roadside	UFPs (NanoScan)	Universal kriging	Exposure model performance, health estimates	1, 3	4
Mobile roadside	UFPs (NanoScan)	Universal kriging	Health estimates corrected for exposure measurement error	3	5
Mobile roadside	UFPs (P-Trak)	Universal kriging	Costs and logistical requirements	4	9
Mobile roadside	UFPs (P-Trak)	Universal kriging and machine learning	Exposure model performance, health estimates	2, 3	8
Mobile roadside	UFPs (NanoScan, P-Trak, and DiSCmini), BC, NO_2, CO_2, $PM_{2.5}$	Universal kriging, machine learning, and multipollutant	Model performance	2	8
Mobile on-road	UFPs (P-Trak)	Universal kriging	Exposure model performance, health estimates	1, 3	6
Low-cost sensors	$PM_{2.5}$ and NO_2	Spatiotemporal	Exposure model performance, health estimates	1, 3	7

[a] The UFP instruments are TSI NanoScan 3910, TSI P-Trak 8525, and Testo DiSCmini.



Commentary Figure 1. Schematic overview of the study design.

detailed air pollution and cognitive function data available at baseline (1994 or later) from the Adult Changes in Thought (ACT) study in Seattle, which is a cohort study of about 5,400 individuals 65 years of age or older (Box 1).

AIR POLLUTION MONITORING

Mobile Monitoring Campaign of UFPs

UFPs were measured with three different instruments: a P-Trak, a NanoScan, and a DiSCmini. All instruments measure particle number concentration but differ in the measurement principle and size ranges captured. Different instruments were used for the analyses across the chapters (Commentary Table 2).

The investigators leveraged real-time mobile monitoring data of UFPs that were collected along nine predefined routes during a year-long campaign from March 2019 to March 2020 using a hybrid vehicle. A total of 309 sites off the side of the road were selected for short-term monitoring along the routes. These sites were intended to represent cohort participants' residential locations, which were on average 611 meters away from their closest roadside measurement site.

Air pollution measurements at the roadside mobile monitoring sites were made for 2 minutes while the vehicle was parked (visits). Visits were repeated on average 30 times throughout the monitoring campaign, and thus, about an hour of data was collected at each site. To obtain a representative annual average, the monitoring was temporally balanced to ensure that all sites were visited during all four seasons and all days of the week, including weekends, and at both early morning and late evening hours (from 5 a.m. to 11 p.m.).

The median of the real-time data was calculated for each visit. Then the investigators transformed the data to reduce the influence of possible extreme values by winsorizing. They winsorized the data across all visits for each site, which replaces values at the tails of the distribution with a fixed percentile value. Values below the 5th and above the 95th percentile were replaced with the values at those thresholds. The winsorized data were then averaged over all visits, log-transformed, and used for subsequent exposure modeling. The mobile roadside monitoring data were used in most of the analyses in the report (Chapters 4, 5, 8, and 9) to investigate their influence on exposure, measurement error, costs, and health estimates.

The investigators also conducted on-road measurements between roadside sites along the nine predefined routes, totaling 1,069 km in the 1,200 km^2 study area. The on-road measurements were aggregated to 100-meter road segments for a total of 5,878 road segments. Each segment was visited an average of 28 times. The investigators excluded some road segments to represent long-term residential exposure better, such as interstates and highways with restricted access, and road segments with fewer than 23 repeat visits.

Box 1: What Is the ACT Study?

The ACT study is a population-based cohort of older cognitively unimpaired adults (65 years or older) in the greater Seattle area. The goal of ACT is to study factors that affect brain aging. Recruitment for the study began in 1994, with a second recruitment wave in 2000, followed by continuous enrollment starting in 2005. Participants are randomly selected from Seattle area members of Group Health Cooperative (now Kaiser Permanente), which is a health maintenance organization. Participants are evaluated every 2 years, where a wealth of data is collected about participants' medical care, physical health, and cognitive health, including their memory and thinking abilities. For the current study, 5,409 participants were included with available information on cognitive function at baseline.

Cognitive function was measured with the Cognitive Abilities Screening Instrument (CASI), which is a 40-item cognitive test that assesses a broad range of cognitive domains (e.g., attention, memory) using scores ranging from 0 to 100. CASI scores were transformed using Item Response Theory (IRT), allowing for a more nuanced interpretation of cognitive function compared to a simple raw total score on the CASI test. For the air pollution analysis, the scores were normalized with values less than 0 indicating lower cognitive function and scores above 0 indicating greater cognitive function than average.

From 2017 to 2020, the ACT study added air pollution measurements from a mobile monitoring campaign and a low-cost fixed-site monitoring campaign specifically designed to represent outdoor exposures for the ACT cohort, which was used in the current study (ACT Air Pollution study).

For more information on the ACT study, go to *https://www.actagingresearch.org/index.php*.

The median of the real-time data was calculated for each 100-meter segment (equivalent to about 10 seconds of observations per visit). Similar to the mobile roadside data, the on-road data were winsorized at the segment level before averaging, and then log-transformed after averaging over all sampling days. On-road UFP data were further adjusted in some analyses to reduce the influence of localized on-road plumes to better represent concentrations at residential locations. Principal components analysis was used to partition simultaneous on-road measurements of multiple pollutants and identify those pollutants associated with localized plumes. An absolute principal component scores model was then used to adjust the UFP data in an iterative process that also leveraged the mobile roadside measurements, as described in detail in Doubleday and colleagues (2023). The mobile on-road monitoring data were used in Chapter 6 of the report to investigate their influence on exposure and health estimates.

Low-Cost Sensor Data for $PM_{2.5}$ and NO_2

The investigators leveraged outdoor $PM_{2.5}$ and NO_2 data from a low-cost sensor campaign conducted in 2017–2020 at about 115 fixed sites, most of them at ACT study participants' residences. $PM_{2.5}$ and NO_2 were measured with a Plantower sensor (PMS A003) and an Alphasense sensor (B43F), respectively. Detailed quality control was conducted. The low-cost sensor data were calibrated daily, using regression models from co-located regulatory measurements, which are described in detail elsewhere (Zuidema et al. 2021; Zusman et al. 2020).

Low-cost sensor data were used in models (see below) that combined them with data from regulatory monitors and other research-grade monitors from many additional locations from 1978 (1996 for NO_2) to 2020 to explore the added value of low-cost sensor data and its design features. In total, $PM_{2.5}$ and NO_2 data from those other data sources were available at 80 and 375 additional locations, respectively. In particular, a snapshot campaign of two or three 2-week measurements at 110 roadside locations conducted in 2018–2019 added substantial spatial information to the models for NO_2.

The investigators averaged all data into 2-week averages. The data were highly unbalanced, with some locations providing longer time series of data, and with only one or a few 2-week observations in other locations. The low-cost sensor data were used in Chapter 7 of the report to investigate their influence on exposure and health estimates.

AIR POLLUTION MODELING

Universal Kriging and Spatiotemporal Models

The investigators subsampled mobile monitoring data to evaluate various exposure assessment study design features, such as fewer visits per site, fewer days of the week, restricted hours of the day, and fewer seasons. The investigators evaluated the impact of those reduced features for both roadside and on-road mobile data. In addition, for the on-road data, they also tested a temporal adjustment for unbalanced diurnal sampling using a fixed-site background monitor and employed a plume correction method to reduce the influence of localized on-road plumes. Each alternative design was resampled 30 times, totaling 480 different design options that were tested for the primary analyses using mobile roadside data.

The investigators used either the full dataset or subsets of measurements to develop annual average exposure estimates using universal kriging models with dimension reduction using partial least squares (for brevity: universal kriging models). Separate models were developed for roadside and on-road mobile data. A suite of geographical covariates (about 200 variables) was available for inclusion in the models, including indicators of land use, roadway proximity, and population density.

Similarly, the investigators subsampled low-cost sensor data to evaluate designs with varying sampling durations, repeated measurements across periods, and sampled locations. Using the subsets of data, they developed up to 20 different spatiotemporal $PM_{2.5}$ models that were designed to accommodate highly imbalanced data. Note that all spatiotemporal models included all data from the regulatory monitors and other research-grade monitors.

They assessed the performance of each model using cross-validation (mobile monitoring data) or a combination of cross-validation and external validation (low-cost sensor data). They reported several measures of performance, including the root mean squared error (RMSE) and the MSE-based explained variance (MSE-R^2). The latter was used instead of the more common regression-based R^2 because it evaluates whether predictions and observations are the same (i.e., are near the one-to-one line), rather than merely correlated. As such, MSE-R^2 assesses both bias and random variation around the one-to-one line. Because of resampling each alternative design, the median of the performance measures and the distribution were reported in the analyses using mobile monitoring data. The models using all data from the mobile roadside campaign or all the low-cost sensor data were taken as reference models.

Machine Learning Models and Other Advanced Statistical Methods

Leveraging all data from the mobile roadside campaign — thus no subsets — and using all pollutants measured (UFPs, BC, $PM_{2.5}$, NO_2, and CO_2), the investigators explored the use of advanced statistical methods beyond the universal kriging model. They developed several machine learning models (six per pollutant), including spatial random forest, that allowed nonlinear and highly complex relationships between geographical predictor variables to be incorporated. They compared the performance of the different models to the earlier-mentioned universal kriging model, which was used as the reference model.

Moreover, because machine learning models are often considered a "black box," they developed a new "variable importance metric" in this study to aid in selecting and interpreting the models. They demonstrated the utility of this metric with the spatial random forest model.

All models developed up to this point were single-pollutant models. In one section of the report, the investigators describe the development and application of a spatial multipollutant model of multiple properties of UFPs (from all three instruments), BC, and NO_2. The model uses a new principal components-based dimension reduction algorithm that considers the spatial patterns in the data while optimizing extrapolation to locations without measurements. The report presents maps of the first three principal component scores for the Seattle area, which highlight the airport and major roads as multipollutant sources. The use of this approach for health estimates analyses of mixtures is identified as future work and is not further summarized in the Commentary.

Those advanced methods were used in one chapter of the report (Chapter 8).

HEALTH ESTIMATES

Each model, except for the multipollutant model, was used to predict the 5-year average UFPs or other pollutant exposures prior to the cognitive function measurement that was obtained at baseline (1994 or later). This exposure was assigned at the residential address level and accounted for residential mobility over the 5 years prior to baseline. The investigators then assessed the association between the 5-year average exposure from each exposure model and baseline cognitive function using standard linear regression. Each model was adjusted for participant age, sex, education, and calendar year (confounder model 1), and associations were expressed per interquartile range of exposure. Results presented in Chapters 4 (mobile roadside data for UFPs) and 6 (mobile on-road data for UFPs) were also adjusted for participant race and socioeconomic status (confounder model 2).

The adjusted associations using the all data exposure model from the mobile roadside campaign for UFPs or from the low-cost sensor data for $PM_{2.5}$ and NO_2 were taken as the reference models. To add context, the associations for cognitive function from the reference models were also expressed as the equivalent of cognitive decline due to aging for a certain number of months.

EXPOSURE MEASUREMENT ERROR ADJUSTMENT

The universal kriging model using all data from the mobile roadside campaign was taken as the reference model in many of the analyses, assuming that those estimates were the best representation of the true annual average. However, even this model contains exposure measurement errors that are, for example, related to the number of locations and times sampled or related to the level of smoothing when fitting the data. Hence, the investigators applied their previously developed bootstrap methods (e.g., Bergen et al. 2016; Szpiro et al. 2011a, 2011b) to quantify the exposure measurement error in the reference model for UFPs, and to correct the health effect estimates accordingly. A key assumption in the measurement error approach — spatial "compatibility" of monitoring sites and cohort locations — was met because the mobile monitoring campaign was specifically designed to capture exposures for the ACT cohort.

Note that the exposure measurement error was not quantified for the UFP models using subsets of mobile monitoring data or for $PM_{2.5}$ and NO_2 models using low-cost sensors.

COSTS AND LOGISTICAL REQUIREMENTS

Using similar exposure design features and universal kriging models as described earlier, the investigators explored the trade-offs between exposure model performance and logistical features (both cost and time) to identify optimal monitoring designs. The investigators used the data from the mobile roadside campaign for those comparisons, with the UFP comparisons only considering the P-Track instrument data. They conducted similar cost–performance analyses using $PM_{2.5}$ and NO_2 low-cost sensor data, but this analysis was described mostly in the Additional Materials and hence not summarized in this Commentary.

They included both up-front and per-drive day costs informed by their own monitoring experience and expenditures for the ACT Air Pollution study. Examples of up-front costs include the purchase of the P-Trak instrument, software development, and various preparation efforts. Most of these costs did not vary by monitoring design. The per-drive-day cost included the cost of staff time for driving, vehicle use, and data management. They expressed the cost as the number of working days, assuming fixed staff costs. They also explored the addition of multiple instruments and pollutants, as well as the addition of a premium for staff working evening and weekend hours. Note that the monetary analysis did not consider costs related to data management and analysis related to exposure model development, which is commonly performed after the monitoring campaign has been completed.

SUMMARY OF RESULTS

MOBILE MONITORING DATA

Performance of Exposure Models Using All Data

The universal kriging reference model using all UFP data from the mobile roadside campaign had a cross-validated MSE-R^2 value of 0.65 (NanoScan) and 0.77 (P-Trak). The cross-validated MSE-R^2 values were 0.65, 0.76, and 0.77 for BC, $PM_{2.5}$, and NO_2, respectively.

The universal kriging model performed slightly better — up to an increase of 0.10 in MSE-R^2 — than the various machine learning models for all pollutants except for BC, where similar performances were observed.

Performance of Exposure Models Using Subsets of Mobile Monitoring Data

The universal kriging models with restricted mobile roadside sampling of UFPs almost always produced lower-performing exposure models compared to the reference model. Models based on datasets with sampling restricted to only one season, sampling conducted only during business hours, and those with few visits to high variability sites had the lowest performance (MSE-R^2 of 0.43–0.48) (**Commentary Table 3**).

Predictions from most designs were highly correlated with predictions from the all data campaign (median Pearson correlations [R] > 0.85), although predictions from the business hour design were consistently less correlated than those from all other designs (R = 0.77).

Commentary Table 3. Performance of Various Exposure Assessment Study Design Features Using Mobile Roadside UFP Data and Its Impact on the Estimated Association Between UFPs and Cognitive Function Using Confounder Model 1

Design Choice	Performance of Exposure Models[a]	% Attenuation of the Association[b]
All-Data Reference Model	*0.65*	*Reference Association: –0.020 per 1,900 Particles/cm³*
Fewer visits		
12	0.59	5%
6	0.53	5%
4	0.51	10%
Fewer seasons for 12 visits in total		
4	0.61	0%
3	0.58	5%
2	0.58	15%
1	0.43	15%
Fewer hours		
Business hours	0.45	60%
Rush hours	0.56	40%
Spatial balance		
Balanced	0.61	0%
Low number of sites with high variability	0.48	10%
High number of sites with high variability	0.59	10%

[a] Median of cross-validated MSE-R^2 from universal kriging models using UFPs (NanoScan) from the mobile roadside campaign.

[b] % attenuation based on the median values, with associations adjusted for participant age, sex, education, and calendar year (confounder model 1).

Using mobile on-road UFP data, the investigators found that most comparisons identified the same design features or elements as important, although with a few notable differences. Spatial balance had a minimal impact on exposure model performances in the on-road models. Strikingly, on-road modeling results were generally similar when road segments were tested 4 times versus 12 times, although the latter produced more stable results.

Comparison of Health Estimates Using Different Exposure Estimates

Exposure of UFPs estimated using the reference model was negatively (adversely) associated with cognitive function at baseline when adjusted for participant age, sex, education, and calendar year (confounder model 1). The UFP association was –0.020 (95% CI: –0.036 to –0.004) per increase of 1,900 particles/cm^3. This estimate is equivalent to accelerated aging of 7.5 months (on average) for cohort participants.

The reduced-sampling designs led to similar findings in terms of negative (adverse) associations between UFP exposure and cognitive function at baseline. However, the strength (magnitude) of the observed negative associations sometimes differed substantially, especially for the business and rush hours designs, which attenuated associations by up to 60% (Commentary Table 3).

Notably, the observed negative association in the reference model disappeared in health models that also adjusted results for race and socioeconomic status (confounder model 2). The null findings from the reference model using confounder model 2 hampered the assessment of the influence of sampling design on health estimates using different exposure estimates for UFPs. Hence, for the aims of their project, the investigators decided to focus on the findings from confounder model 1 throughout the report.

Exposure Measurement Error Adjustment

The observed negative associations using confounder model 1 were affected more by features of the mobile monitoring design than by accounting for exposure measurement error in the reference exposure model. The investigators reported only a modest influence on the observed negative associations when results were adjusted for exposure measurement error in the reference model (6% on the association, 13% on the confidence intervals).

Costs and Logistical Requirements

The investigators found that a mobile monitoring study with roadside sampling of UFPs with at least 12 visits per location optimized exposure model performance while also limiting costs. Furthermore, the investigators noted that it is important that the mobile monitoring campaign covers all days of the week, most hours of the day (including early morning and late evening hours), and at least two seasons.

LOW-COST SENSOR DATA

The addition of low-cost sensor data improved $PM_{2.5}$ exposure modeling. The spatiotemporal model using all data had a cross-validated MSE-R^2 value of 0.84, whereas the model with only the regulatory monitors and other research-grade monitors had a value of 0.69. Furthermore, increasing the number of low-cost sensor locations and repeated measurements resulted in better exposure model performance.

In contrast, in most comparisons, the addition of low-cost sensor data improved the NO_2 estimates only slightly in models that combined data from regulatory monitors and other research-grade monitors. For example, the cross-validated MSE-R^2 value increased from 0.82 to 0.85 in a spatial comparison. Reasons may relate to the large amount of spatial information already in the model from many additional locations (375 in total), using other data sources that were not available for $PM_{2.5.}$

Largely null findings were reported between $PM_{2.5}$, NO_2, and cognitive function for the various spatiotemporal exposure models with and without low-cost sensor data using confounder model 1. The null results thus hampered the assessment of the influence of adding low-cost sensor data for health effect estimates.

HEI IMPROVED EXPOSURE ASSESSMENT STUDIES REVIEW PANEL'S EVALUATION

In its independent review, the HEI Review Panel thought the study was well-motivated and appreciated that it leveraged detailed air pollution and cognitive function data from the ACT study in Seattle. They thought the study was comprehensive, with thorough analyses and findings that will be of broad interest and value to a wide audience.

STRENGTHS OF THE STUDY

The Panel noted several strengths of the research. First, the Panel recognizes the benefits of an extensive year-long mobile monitoring campaign that includes both roadside and on-road sampling. The investigators leveraged a rich dataset on UFPs and other pollutants that covered various times of day between 5 a.m. and 11 p.m., weekdays, and weekends — thus including those times of day when people might be more likely to be at home — and all four seasons. Many other mobile monitoring campaigns have collected less data, sampled during more restricted periods, such as business hours only, or had short monitoring durations lasting only a few weeks or months. Some of those studies have used continuous measurements at a fixed reference site to account for temporal variation (Kim et al. 2023).

Second, using this detailed dataset, the investigators evaluated various exposure assessment study design features and provided practical guidance on future mobile monitoring campaigns. Developing this guidance addressed a clear research gap and should be of interest to a wide audience. The consideration of the study design costs in developing the guidance, as informed by their experience, was also appreciated.

Third, the extensive air pollution exposure modeling and rigorous evaluation of their performance are strengths of the study. The investigators developed a suite of models, including universal kriging, spatiotemporal models, machine learning, and other advanced statistical models. The large number of geographical covariates (about 200 variables) available for inclusion in the models was notable. The development of a novel variable importance metric that is applicable to machine learning methods such as spatial random forest may be an important contribution. Moreover, the investigators reported several measures of performance to test accuracy and possible bias — thereby providing an in-depth performance assessment.

Fourth, the Panel found that the analysis to adjust for exposure measurement error in the health analyses was a valuable contribution. Accounting for the inherent (spatially varying) uncertainty and biases in modeled estimates of air pollution remains largely an unresolved problem in air pollution epidemiology (Samoli and Butland 2017; Sheppard et al. 2012), and advances in this area are much needed.

Fifth, the Panel was impressed by the large number of publications resulting from the work, as nicely documented in the report's annotated bibliography.

Although the Panel broadly agreed with the investigators' conclusions, some limitations should be considered when interpreting the results, as explained next.

FOCUS ON UFPs AND USE OF DIFFERENT INSTRUMENTS

Although the HEI-funded work encompassed analyses of multiple pollutants, most of the report is focused on UFPs from mobile monitoring or $PM_{2.5}$ and NO_2 from low-cost monitoring. Few comparisons across pollutants are included in the main report, although more information is presented in the Additional Materials and other publications. The Panel thought this limited the generalizability of the findings and that additional research is warranted for the other pollutants.

For the evaluation of exposure design features using mobile roadside and mobile on-road monitoring data of UFPs, findings were presented in different stand-alone chapters, and different instruments were selected (a NanoScan and a P-Trak) for various analyses. This difference makes a direct comparison between the two monitoring approaches difficult.

Those instruments differ in the size range captured: 10–420 nm for the NanoScan and 20–1,000 nm for the P-Trak. The lower cut-off measurement is usually critical because most UFPs are smaller than 20 nm and not captured by the P-Trak (HEI Review Panel on Ultrafine Particles 2013). Even small differences in the lower cut point in the range below 20 nm can lead to substantial differences in the particle number concentration. The very small particles (< 20 nm) are also the particles that might exhibit the highest variability in space and time (HEI Review Panel on Ultrafine Particles 2013). Indeed, in the current study, the NanoScan measured concentrations that were roughly 50% higher than those of the P-Trak, with more variability (contrast).

The investigators did not report detailed information on particle size distributions, preventing an in-depth particle size distribution analysis. However, the investigators conducted sensitivity analyses using P-Trak for the mobile roadside analyses and reported similar results, alleviating the concern to some extent. Note that they could not conduct a similar sensitivity analysis using the NanoScan for the mobile on-road analyses because of the time resolution (60 seconds for NanoScan versus 1 second for the P-Trak). The investigators also added a qualitative comparison section of the two monitoring approaches in the synthesis chapter, based on the Panel's recommendation, which was appreciated.

REMOVING THE INFLUENCE OF POSSIBLE EXTREME VALUES

The investigators calculated medians for each visit (or segment), winsorized across all visits for each site, and then log-transformed the data for subsequent exposure modeling. All three approaches are meant to reduce the influence of possible extreme values and increase evenness in the data. A sensitivity analysis to investigate how each approach — in particular winsorizing the mobile monitoring data — would affect the exposure models was missing from the report. A more thorough examination of the influence of possible extreme values was, however, included in an earlier paper (Blanco et al. 2022).

The Panel concluded that winsorizing the data improved exposure model performance for some pollutants (e.g., BC, NO_2, $PM_{2.5}$) but not consistently for UFPs, which was the focus of most of the report. For example, in Blanco and colleagues (2022), the "median of medians" approach had improved out-of-sample model performance in terms of lower RMSE and near identical MSE-R^2 values compared to winsorizing, particularly for the NanoScan and P-Trak instruments used in Chapters 4 and 6. The investigators justified their approach to also winsorize UFPs for "consistency across pollutants," but as it turned out, only limited comparisons across pollutants were included in the report.

THE HEALTH ANALYSES WERE CONSIDERED LIMITED

The Panel members thought the study's main strengths lie in its contributions to methodological development regarding improved exposure assessment design rather than the evaluation of health estimates. Although the Panel appreciated the complexities involved, the health analyses were considered limited for three reasons.

First, most exposure models used were based on measurements conducted up to 25 years after the health outcome. The Panel thought the investigators could have used health outcome data from later years to better align with the 2019–2020 exposure models from the ACT study, which is an ongoing cohort with participants who are followed up every 2 years. This temporal mismatch between the period captured by the mobile measurements and the exposure window most relevant for epidemiological purposes is also apparent in some other cohort studies investigating UFPs (e.g., Alexeef et al. 2018; Bai et al. 2019; Downward et al. 2018; Pond et al. 2022; Weichenthal et al. 2017, 2024). However, the temporal mismatch in other studies is typically shorter (e.g., up to 10 years after the end of follow-up), and some of those studies (e.g., Weichenthal et al. 2024) applied a back-casting procedure using supplemental data to partially overcome the lack of air pollution data in earlier years. The investigators did not apply a back-casting procedure, although one was originally proposed, and assumed that air pollution exposure surfaces remained constant over all that time, which is a large assumption that they did not test in the report. The Panel thought this assumption is difficult to defend because air pollution concentrations have been declining over the past few decades in many high-income countries, due largely to successful air quality regulation and subsequent reductions in emissions from major air pollution sources, including transportation and power generation (Boogaard et al. 2024; US EPA 2016, 2022; WHO 2021).

Second, the health analysis was a cross-sectional analysis of one measure of cognitive function. In the original application, the investigators planned to use longitudinal data on cognitive decline and dementia incidence from the ACT cohort. Because of delays in accessing the health data and because the models they developed were more computationally intensive than expected, the investigators decided to conduct the current health analysis within the scope of the current project. The Panel would welcome the more advanced health analyses that the investigators are planning, as alluded to in the report.

Third, the Panel was concerned that residual confounding was likely in the analyses (confounder model 1) due to inadequate adjustment for characteristics that are correlated with air pollution and cognitive function, most notably socioeconomic status. Findings differed for models that adjusted for race and socioeconomic status (confounder model 2) compared to those that did not (confounder model 1), as documented in Chapters 4 and 6 of the report. The Panel thought the authors should have adjusted for socioeconomic status in all health analyses. The Panel also noted the lack of information on potential individual lifestyle covariates, such as smoking.

For those reasons, the investigators avoided the use of causal inference language in the report, as supported by the Panel. Nevertheless, the Panel recommended caution when interpreting the findings of the health analyses.

USE OF REAL-WORLD DATA VERSUS SIMULATIONS

The Panel thought simulations would have complemented the study because of the limitations in the health analyses using real-world data. In particular, the null findings from the reference exposure model using confounder model 2 hampered the assessment of the influence of sampling design on health estimates using different exposure estimates. An advantage of using simulated data is that the underlying "true" health effects are known in that scenario, and one can systematically test one feature while holding all other conditions constant. The challenge with simulation studies is that they might not adequately represent the real world. Some carefully designed simulations, along the lines discussed in the report, could have shed light on the differing health results between confounder models. However, this would have increased the scope of the project beyond what the investigators originally proposed.

GENERALIZABILITY OF GUIDANCE ON MOBILE MONITORING CAMPAIGNS

The investigators provided practical guidance for the implementation of future mobile monitoring campaigns. However, the Panel had some concerns about the generalizability of the findings related to the improved exposure assessment design. The air pollution exposure estimates in the analyses were relatively low, and the variability (contrast) was limited. The average UFP concentrations were low (10,000 particles/cm^3 measured with the NanoScan), typical of urban background areas in North America, and lower than typical near-roadway locations. Concentrations of the other pollutants were similarly low; for example, the average $PM_{2.5}$ concentration measured using low-cost sensors was 6 μg/m^3, and the interquartile range was 1 μg/m^3. Also, Seattle has a temperate climate characterized by moderate temperatures with mild winters and dry summers with little extreme heat or cold.

The performance of the different reduced-sampling models was compared against the reference model, which included all roadside monitoring data. Results and recommendations regarding adequate or optimal numbers of sites and visits might be specific to the Seattle area and the time period of the study. Relatedly, it is important to mention that the mobile monitoring campaign was specifically designed to capture exposures for the ACT cohort (in other words, the monitoring sites and cohort locations were spatially "compatible"). In other studies, residential locations of interest might not be known in advance, so monitoring routes might need to be selected based on different criteria. Earlier campaigns have often selected monitoring locations based on maximizing air pollution variation by including different geographic features, land uses, or different sources of air pollution (Kim et al. 2023).

Thus, some caution is warranted in generalizing the findings, and further research in other cities would be helpful to assess the generalizability of the specific findings related to exposure assessment design. Note that the findings were not affected by the COVID-19 pandemic in 2020 because monitoring was completed before that time.

COMPARISON OF MOBILE MONITORING GUIDANCE WITH OTHER STUDIES

Regarding the guidance on future mobile monitoring campaigns, the number of repeated measurements (12 visits) per location that the investigators considered optimal aligns with an intensive mobile on-road monitoring study using Google Street View cars in Oakland, CA (Apte et al. 2017). Apte and colleagues (2017) documented that up to about 10 repeated driving days, the stability of the measured average concentration of BC, NO, and NO_2 increased, and after about 20 driving days, the stability of the average did not improve appreciably with further repeats (Apte et al. 2017). Note that UFPs were not measured in the Oakland study.

The investigators further emphasized the importance of including sampling beyond business hours, including extended times of the day and weekends. However, a recent analysis, which is part of the HEI report by Gerard Hoek and colleagues funded under the same RFA as the current study, did not identify the time of day as an important feature that would explain some of the heterogeneity of effect estimates observed (Hoek et al. 2025). Among many other exposure assessment approaches, Hoek and colleagues compared three mobile monitoring studies (Kerckhoffs et al. 2016, 2017, 2021) that excluded rush hours with a study using Google Street View cars that monitored from 8 a.m. to 10 p.m. on weekdays in the Netherlands (Kerckhoffs et al. 2022). However, they did not have access to a mobile monitoring study that covered 24 hours of the day and weekend days, which may partly explain the difference in findings from these two studies.

Furthermore, Sheppard and colleagues noted that it would be important that the exposure sampling in mobile monitoring campaigns include at least two seasons. Although the Panel generally agreed with the seasonal sampling requirement, the importance of season might be characterized by patterns in temperature, sunlight, humidity, precipitation, and wind (Pérez et al. 2020). Hence, the Panel emphasized that "seasons" depend on geographical location, and how many "seasons" exist varies by region.

SUMMARY AND CONCLUSION

The study by Dr. Sheppard and colleagues compared the performance of different exposure assessment study design features on long-term exposure estimates in Seattle, Washington. The investigators determined the impact on exposure prediction and health effect estimates using various approaches to sample data collected from an earlier mobile monitoring campaign and a fixed-site monitoring campaign with low-cost sensors. The investigators used either the full dataset or subsets of measurements to develop annual average universal kriging models, machine learning models, and other advanced statistical models. The study leveraged detailed air pollution data and cognitive function data at baseline (1994) from the Adult Changes in Thought (ACT) Air Pollution study in Seattle, a cohort study of older adults.

The study provides practical guidance on future mobile monitoring campaigns, which addresses a clear research gap. The extensive year-long mobile monitoring campaign and the evaluation of various exposure assessment study design features were strengths of the study. Another strength was the extensive air pollution exposure modeling and rigorous evaluation of their performance. The Panel was also impressed by the large number of publications resulting from the work.

The investigators found that a mobile monitoring study with roadside sampling of UFPs with at least 12 visits per location optimized exposure model performance while limiting costs. Furthermore, the investigators noted that it is important that the exposure sampling in mobile monitoring campaigns covers all days of the week, most hours of the day (including early morning and late evening hours), and at least two seasons.

Although the Panel broadly agreed with the investigators' conclusions, some limitations should be considered when interpreting the results. Many of the analyses focused on UFPs, and there were few comparisons across pollutants, albeit more information is presented in the Additional Materials and other publications. For the evaluation of exposure design features using mobile roadside and mobile on-road monitoring data of UFPs, findings were presented in different stand-alone chapters, and different monitoring instruments were selected for various analyses. This makes a direct comparison between the two monitoring approaches difficult. An analysis investigating how the removal of possible extreme values affected the subsequent exposure models was missing from the report, but was included in a paper resulting from this work.

The Panel recommended caution in interpreting the findings from the health analyses and thought some carefully designed simulations would have complemented the real-world health study. The health analyses were considered limited, particularly because most of the exposure models used were based on measurements conducted up to 25 years later than the health outcome. In addition, some analyses lacked adjustment for important confounding variables, most notably socioeconomic status. Further research in other cities and pollutants would be helpful to assess the generalizability of the specific findings related to exposure assessment design.

The comprehensive report includes many findings that will be of broad interest and value to a wide audience.

ACKNOWLEDGMENTS

The HEI Review Committee is grateful to the Improved Exposure Assessment Review Panel for their thorough review of the study. The Committee is also grateful to Allison Patton for oversight of the study, to Hanna Boogaard for assistance with review of the Investigators' Report and preparation of its Commentary, to Tom Zaczkiewicz for editing the Investigators' Report and its Commentary, and to Kristin Eckles for her role in preparing this Research Report for publication.

REFERENCES

Alexeeff SE, Roy A, Shan J, Liu X, Messier K, Apte JS, et al. 2018. High-resolution mapping of traffic-related air pollution with Google street view cars and incidence of cardiovascular events within neighborhoods in Oakland, CA. Environ Health 17:38.

Apte JS, Messier KP, Gani S, Brauer M, Kirchstetter TW, Lunden MM, et al. 2017. High-resolution air pollution mapping with Google Street View cars: exploiting big data. Environ Sci Technol 51:6999–7008.

Apte JS, Chambliss SE, Messier KP, Gani S, Upadhya AR, Kushwaha M, et al. 2024. Scalable multipollutant exposure assessment using routine mobile monitoring platforms. Research Report 216. Boston, MA: Health Effects Institute.

Bai L, Weichenthal S, Kwong JC, Burnett RT, Hatzopoulou M, Jerrett M, et al. 2019. Associations of long-term exposure to ultrafine particles and nitrogen dioxide with increased incidence of congestive heart failure and acute myocardial infarction. Am J Epidemiol 188:151–159.

Bergen S, Sheppard L, Kaufman JD, Szpiro AA. 2016. Multipollutant measurement error in air pollution epidemiology studies arising from predicting exposures with penalized regression splines. J R Stat Soc Ser C-Appl Stat 65:731–753.

Blanco MN, Gassett A, Gould T, Doubleday A, Slager DL, Austin E, et al. 2022. Characterization of annual average traffic-related air pollution concentrations in the Greater Seattle Area from a year-long mobile monitoring campaign. Environ Sci Technol 56:11460–11472.

Boogaard H, Crouse DL, Tanner E, Mantus E, van Erp AM, Vedal S, et al. 2024. Assessing adverse health effects of long-term exposure to low levels of ambient air pollution: The HEI experience and what's next? Environ Sci Technol 58:12767–12783.

Dockery DW, Pope CA 3rd, Xu X, Spengler JD, Ware JH, Fay ME, et al. 1993. An association between air pollution and mortality in six US cities. N Engl J Med 329:1753–1759.

Doubleday A, Blanco MN, Austin E, Marshall JD, Larson TV, Sheppard L. 2023. Characterizing ultrafine particle mobile monitoring data for epidemiology. Environ Sci Technol 57:9538–9547.

Downward GS, van Nunen EJHM, Kerckhoffs J, Vineis P, Brunekreef B, Boer JMA, et al. 2018. Long-term exposure to ultrafine particles and incidence of cardiovascular and cerebrovascular disease in a prospective study of a Dutch cohort. Environ Health Perspect 126:127007.

Hatzopoulou M, Valois MF, Levy I, Mihele C, Lu G, Bagg S, et al. 2017. Robustness of land-use regression models developed from mobile air pollutant measurements. Environ Sci Technol 51:3938–3947.

HEI 2010. Traffic-Related Air Pollution: A Critical Review of the Literature on Emissions, Exposure, and Health Effects. Special Report 17. Boston, MA: Health Effects Institute.

HEI Review Panel on Ultrafine Particles. 2013. Understanding the health effects of ambient ultrafine particles. Perspectives 3. Boston, MA: Health Effects Institute.

HEI. 2022. Systematic Review and Meta-analysis of Selected Health Effects of Long-Term Exposure to Traffic-Related Air Pollution. Special Report 23. Boston, MA: Health Effects Institute.

Hoek G. 2017. Methods for assessing long-term exposures to outdoor air pollutants. Curr Environ Health Rep 4:450–462.

Hoek G, Bouma F, Janssen N, Wesseling J, van Ratingen S, Kerckhoffs J, et al. 2025. Comparison of long-term air pollution exposure from mobile and routine monitoring, low-cost sensors, and dispersion models. Research Report 226. Boston, MA: Health Effects Institute.

IARC (International Agency for Research on Cancer) Working Group on the Evaluation of Carcinogenic Risks to Humans. 2016. Outdoor air pollution. IARC Monogr Eval Carcinog Risks Hum 109:9–444.

Kerckhoffs J, Hoek G, Messier KP, Brunekreef B, Meliefste K, Klompmaker JO, et al. 2016 Comparison of ultrafine particle and black carbon concentration predictions from a mobile and short-term stationary land-use regression model. Environ Sci Technol 50:12894–12902.

Kerckhoffs J, Hoek G, Vlaanderen J, van Nunen E, Messier K, Brunekreef B, et al. 2017. Robustness of intraurban land-use regression models for ultrafine particles and black carbon based on mobile monitoring. Environ Res 159:500–508.

Kerckhoffs J, Hoek G, Gehring U, Vermeulen R. 2021. Modelling nationwide spatial variation of ultrafine particles based on mobile monitoring. Environ Int 154:106569.

Kerckhoffs J, Khan J, Hoek G, Yuan Z, Hertel O, Ketzel M, et al. 2022. Hyperlocal variation of nitrogen dioxide, black carbon, and ultrafine particles measured with Google Street View cars in Amsterdam and Copenhagen. Environ Int 170:107575.

Kim SY, Blanco MN, Bi J, Larson TV, Sheppard L. 2023. Exposure assessment for air pollution epidemiology: A scoping review of emerging monitoring platforms and designs. Environ Res 223:115451.

Larkin A, Hystad P. 2017. Towards personal exposures: how technology is changing air pollution and health research. Curr Environ Health Rep 4:463–471.

Morawska L, Thai PK, Liu X, Asumadu-Sakyi A, Ayoko G, Bartonova A, et al. 2018. Applications of low-cost sensing technologies for air quality monitoring and exposure assessment: How far have they gone? Environ Int 116:286–299.

Messier KP, Chambliss SE, Gani S, Alvarez R, Brauer M, Choi JJ, et al. 2018. Mapping air pollution with Google street view cars: efficient approaches with mobile monitoring and land use regression. Environ Sci Technol 52:12563–12572.

Pérez IA, García MÁ, Sánchez ML, Pardo N, Fernández-Duque B. 2020. Key points in air pollution meteorology. Int J Environ Res Public Health 17:8349.

Pond ZA, Saha PK, Coleman CJ, Presto AA, Robinson AL, Pope CA III. 2022. Mortality risk and long-term exposure to ultrafine particles and primary fine particle components in a national US Cohort. Environ Int 167:107439.

Pope CA 3rd, Burnett RT, Thun MJ, Calle EE, Krewski D, Ito K, et al. 2002. Lung cancer, cardiopulmonary mortality, and long-term exposure to fine particulate air pollution. JAMA 287:1132–1141.

Roque NA, Andrews H, Santos-Lozada AR. 2025. Identifying air quality monitoring deserts in the United States. Proc Natl Acad Sci U S A 122:e2425310122.

Samoli E, Butland BK. 2017. Incorporating measurement error from modeled air pollution exposures into epidemiological analyses. Curr Environ Health Rep 4:472–480.

Sheppard L, Burnett RT, Szpiro AA, Kim SY, Jerrett M, Pope CA 3rd, et al. 2012. Confounding and exposure measurement error in air pollution epidemiology. Air Qual Atmos Health 5:203–216.

Szpiro AA, Paciorek CJ, Sheppard L. 2011. Does more accurate exposure prediction necessarily improve health effect estimates? Epidemiology 22:680–685.

Szpiro AA, Sheppard L, Lumley T. 2011. Efficient measurement error correction with spatially misaligned data. Biostatistics 12:610–623.

US EPA (United States Environmental Protection Agency). 2016. Integrated science assessment for oxides of nitrogen–health Criteria. EPA/600/R-15/068. Washington, DC: US EPA.

US EPA (United States Environmental Protection Agency), 2019. Integrated science assessment for particulate matter. EPA/600/R-19/188. Washington, DC: US EPA.

US EPA (United States Environmental Protection Agency). 2022. Supplement to the 2019 integrated science assessment for particulate matter. EPA/600/R-22/028. Washington, DC: US EPA.

US EPA (United States Environmental Protection Agency). 2024. Final rule: Reconsideration of the national ambient air quality standards for particulate matter. Available: *https://www.epa.gov/pm-pollution/final-reconsideration-national-ambient-air-quality-standards-particulate-matter-pm* [accessed May 22, 2025].

Weichenthal S, Bai L, Hatzopoulou M, Van Ryswyk K, Kwong JC, Jerrett M, et al. 2017. Long-term exposure to ambient ultrafine particles and respiratory disease incidence in Toronto, Canada: A cohort study. Environ Health 16:64.

Weichenthal S, Lloyd M, Ganji A, Simon L, Xu J, Venuta A, et al. 2024. Long-Term Exposure to Outdoor Ultrafine Particles and Black Carbon and Effects on Mortality in Montreal and Toronto, Canada. Research Report 217. Boston, MA: Health Effects Institute.

WHO (World Health Organization). 2021. WHO Global Air Quality Guidelines: Particulate Matter ($PM_{2.5}$ and PM_{10}), Ozone, Nitrogen Dioxide, Sulfur Dioxide, and Carbon Monoxide. Geneva: World Health Organization.

Zuidema C, Schumacher CS, Austin E, Carvlin G, Larson TV, Spalt EW, et al. 2021. Deployment, calibration, and cross-validation of low-cost electrochemical sensors for carbon monoxide, nitrogen oxides, and ozone for an epidemiological study. Sensors 21:4214.

Zusman M, Schumacher CS, Gassett AJ, Spalt EW, Austin E, Larson TV, et al. 2020. Calibration of low-cost particulate matter sensors: model development for a multi-city epidemiological study. Environ Inter 134:105329.

ABBREVIATIONS AND OTHER TERMS

A1	feature class code road type – interstate highways (primary highway with limited access)
A2	feature class code road type – primary highway without limited access
A3	feature class code road type – secondary and connecting roads
A4	feature class code road type – local, neighborhood, and rural roads
ACT	Adult Changes in Thought
ACT-AP	Adult Changes in Thought Air Pollution
AD	Alzheimer's disease
BC	black carbon
CASI	Cognitive Abilities Screening Instrument
CASI-IRT	CASI transformed using item response theory
CO_2	carbon dioxide
CV	cross-validation
CV-R^2	cross-validation based R^2
CV-MSE R^2	cross-validated mean squared error R^2
CV-RMSE	cross-validated root mean squared error
DEEDS	Diesel Exhaust Exposure in the Duwamish Study (Schulte et al. 2015)
DiSCmini	instrument measuring UFP
FEM	Federal Equivalent Method
FRM	Federal Reference Method
HMO	health maintenance organization
IQR	interquartile range
IRB	institutional review board
MAP	MESA Air Pilot (Wilton et al. 2010)
ME	measurement error
MESA Air	Multi-Ethnic Study of Atherosclerosis and Air Pollution
MSE	mean squared error
NanoScan	instrument measuring UFP
NO_2	nitrogen dioxide
P-Trak	instrument measuring UFP
PC	principal component
PCA	principal component analysis
PLS	partial least squares
$PM_{2.5}$	particulate matter ≤2.5 µm in aerodynamic diameter
PNC	particle number concentration in pt/cm^3
PSCAA	Puget Sound Clean Air Agency
PSID	Panel Study on Income Dynamics (Liu et al. 2003)
R^2	coefficient of determination or explained variance
RAD	Remote Air Data
RapPCA	representative and predictive PCA
RF	random forest
RMSE	root mean squared error
SD	standard deviation
SE	standard error
SES	socioeconomic status
SpatRF	spatial random forest
SpatRF-NP	SpatRF optimized using a nonparametric approach
SpatRF-PL	SpatRF optimized using pseudo-likelihood
TPRS	spatial smoothing via thin plate regression splines
TRAP	traffic-related air pollution
UFP	ultrafine particles
UK	universal kriging
UK-PLS	universal kriging followed by PLS
US EPA	United States Environmental Protection Agency
Yesler	Yesler Terrace study (Wong 2010)