Ultrafine particles and late-life cognitive function: Influence of stationary mobile monitoring design on health inferences

Magali N. Blanco,[1] Adam A. Szpiro,[2] Paul K. Crane,[3] Lianne Sheppard[1,2]

[1]Department of Environmental and Occupational Health Sciences, University of Washington, Seattle, WA USA
[2]Department of Biostatistics, University of Washington, Seattle, WA USA
[3]Department of Medicine, University of Washington, Seattle, WA, USA

Contact Author:
Magali N. Blanco
magali@uw.edu
Dept of Environmental and Occupational Health Sciences
Box 351618
University of Washington
Seattle, WA 98195-1618

Conflicts of interest:
The authors declare they have no conflicts of interest related to this work to disclose.

Abstract

Growing evidence links ultrafine particles (UFP) to neurotoxicity, but human studies remain limited. Various mobile monitoring approaches have been used to collect repeated short-term air pollution samples and develop human exposure models. However, whether design choices impact epidemiologic inferences, including for UFP and cognitive function, remains unclear.

We evaluated the adjusted association between UFP number concentration (PNC) and late-life cognitive function (Cognitive Abilities Screening Instrument – Item Response Theory [CASI-IRT]) in the Adult Changes in Thought cohort (N=5,283) by leveraging an extensive roadside mobile monitoring campaign specifically designed for epidemiologic application. To assess the impact of common, reduced monitoring approaches on this association, we repeatedly subsampled UFP measures from the extensive campaign, developed exposure models, and evaluated the degree to which associations were impacted.

In a reduced adjustment model, the mean baseline CASI-IRT score decreased by 0.020 (95% CI: -0.036, -0.004) per 1,900 pt/cm^3 increase in PNC. Associations were consistent across most sampling designs, including fewer visits (median point estimate -0.019, IQR: -0.022, -0.016), fewer seasons (-0.019, IQR: -0.021, -0.016), and unbalanced sampling (-0.018, IQR: -0.022, -0.016), with very unbalanced designs yielding more variable estimates. In the primary adjustment model, the CASI-IRT score increased by 0.002 (95% CI: -0.016, 0.020), with similar estimates across fewer visits (0.002, IQR: -0.001, 0.004), fewer seasons (0.000, IQR: -0.001, 0.003), and unbalanced sampling (0.001, IQR: -0.001, 0.004). Rush hour designs were more similar (0.002, IQR: 0.000, 0.003) than business hour designs (0.006, IQR: 0.005, 0.007), but the opposite was true when temporal adjustments were applied (rush: -0.003, IQR: -0.005, -0.001; business: 0.002, IQR: 0.001, 0.004). We observed similar trends in sensitivity and secondary analyses.

We found no link between UFP exposures and late-life cognition. While various monitoring approaches may be used to capture epidemiologic inferences, extending beyond weekday business and rush hours is likely important.

1 Introduction

A growing body of evidence links ultrafine particles (UFPs), typically defined as particles ≤ 100 nm in diameter, and other traffic-related air pollutants (TRAP) to adverse health effects (Brugge & Fuller, 2020; Flood-Garibay et al., 2023; HEI, 2013; US EPA, 2019). While research on UFPs is still emerging, these have been associated with increased levels of inflammation, oxidative stress, neurotoxicity, and neurodegeneration (HEI, 2013). Only a few studies have explored the link between UFP exposures and late-life cognitive function in humans, and these have produced mixed results. For example, investigators linked elevated residential UFP levels to an increased incidence of malignant brain tumors (Weichenthal et al., 2020). However, other studies, including one on late-life dementia incidence in the Seattle-based Adult Changes in Thought (ACT) study (Blanco et al., 2024) and another on cognitive decline in older adults recruited from Alzheimer's Disease Research Centers (ADRC) throughout the US (Gan et al., 2023), have reported no associations with UFP exposures. Despite the growing interest, UFPs are not routinely monitored by traditional long-term fixed-site networks like other pollutants, such as fine particulate matter ($PM_{2.5}$) or nitrogen dioxide (NO_2), making it challenging to characterize population-level exposures needed for epidemiologic studies.

Mobile monitoring campaigns, which use a mobile platform such as a vehicle to collect repeated short-term air samples from various locations, are commonly implemented to address this gap and capture the high spatial variability of UFPs (Kim et al., 2023). Over the past few decades, these campaigns have provided valuable insights into pollutant sources, spatiotemporal variability, pollution "hotspots," commuter exposures, and more (Apte et al., 2017; Austin et al., 2021; Karumanchi et al., 2021; Knibbs et al., 2011; Weichenthal et al., 2016). However, monitoring designs vary substantially, and there are no standard protocols. Most campaigns collect only a handful of measurements (visits) per site (median: ~4, range: ~1-40), last between a few weeks and approximately 3 months, and sample exclusively during weekday business or rush hours (Kim et al., 2023). Nearly all collect samples in an unbalanced fashion, with some locations receiving more visits than others due to various logistical constraints. Most campaigns produce poor to moderately performing exposure assessment models.

While an increasing number of campaigns aim to develop exposure models for epidemiologic applications, limited attention has been given to the role of monitoring design in this context, despite its potential to bias long-term exposure assessments (Blanco, Doubleday, et al., 2022; Blanco et al., 2023; Blanco, Gassett, et al., 2022; Sheppard & Blanco, Under Review). This setting is unique from other applications since the goal is often to estimate longer-term, around-the-clock exposures for residential locations rather than shorter-term, weekday, on-road exposures.

The objective of this study was to investigate the association between UFPs and late-life cognitive function and to evaluate how emerging stationary (temporary roadside stop) monitoring designs impact exposure assessment models and subsequent health inferences. We leveraged data from the Adult Changes in Thought (ACT) cohort, a large prospective study on the aging brain (Kukull et al., 2002), and an extensive mobile monitoring campaign conducted in the greater Seattle area specifically designed for epidemiologic applications (Blanco, Gassett, et al., 2022). We conducted a subsampling study to evaluate the impact of common monitoring

designs and provide guidance on the design features that can best support epidemiologic studies.

2 Methods

Figure 1 summarizes the analytic method further detailed in this section. We use extensive roadside mobile monitoring measures to develop reference all-data air pollution exposure prediction models, use these to evaluate baseline ACT participant exposures, and evaluate the association between cognitive function and UFPs. To investigate the role of mobile monitoring design, we subsampled the air pollution data following common field sampling approaches and used the resulting data to develop comparable exposure and health inference models. We compared these results to those from the all-data exposure model.

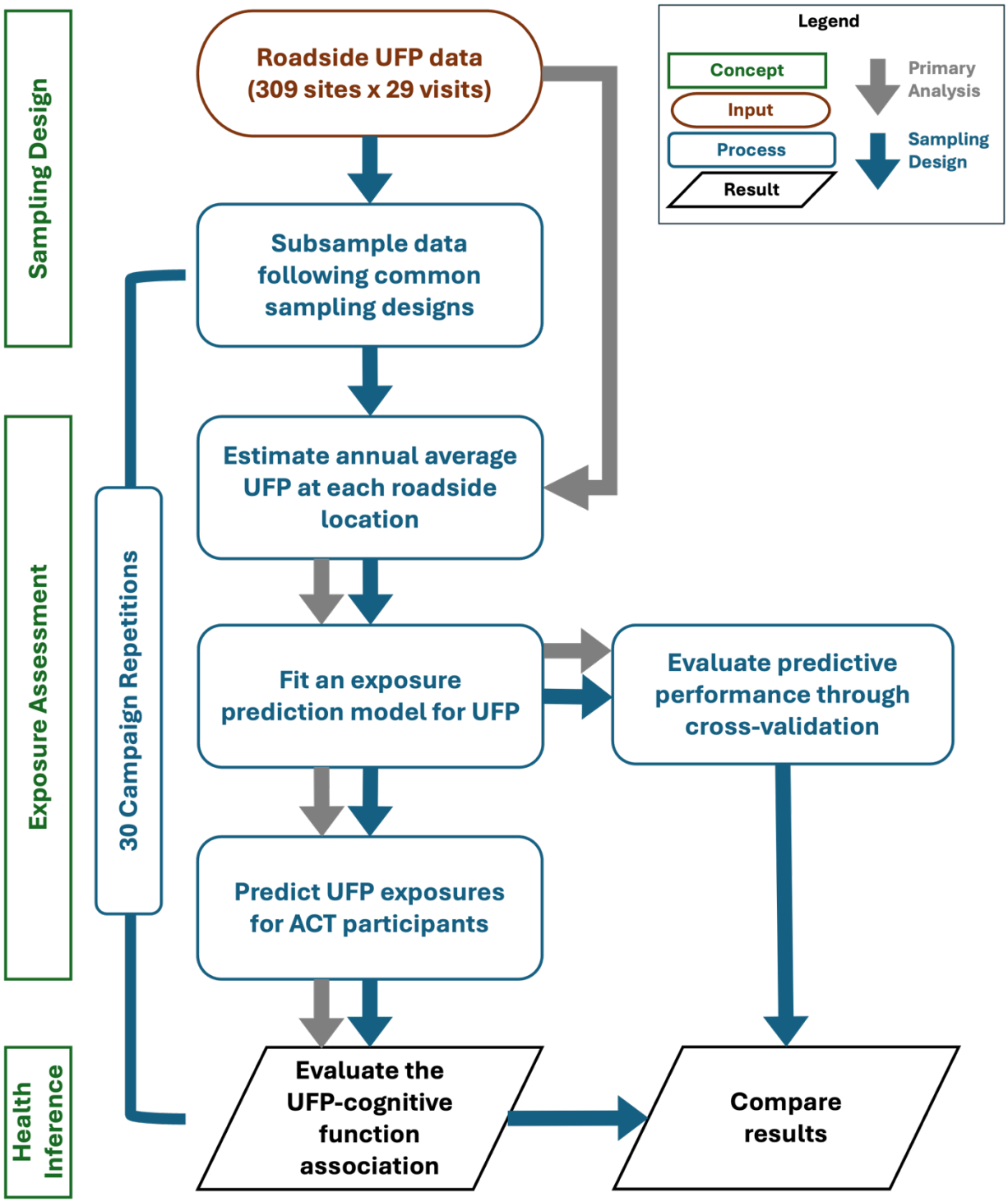

Figure 1. Overview of analytic approach used to evaluate cognitive function, UFP exposures, and the role of air pollution monitoring design.

2.1 The ACT Cohort and Outcome Ascertainment

We evaluated cognitive function in ACT, a community-based, prospective cohort study in the greater Seattle area that has been investigating the aging brain since 1994 (Kukull et al., 2002). The study randomly invites elderly (65+ yr) members of the Kaiser Permanente Washington integrated healthcare delivery system (formerly Group Health Cooperative) to participate. Invitees are assessed for cognitive function at baseline using the Cognitive Abilities Screening Instrument (CASI), which combines common screening tests including the Mini-Mental State Examination (MMSE) and the Hasegawa Dementia Rating Scale to quantitatively assess attention, concentration, orientation, short- and long-term memory, language abilities, judgement, and other functions (Teng et al., 2004). People with high cognitive scores (scores of $\geq$ 86/100) are enrolled. People with low scores are evaluated with a comprehensive neuropsychological battery and focused neurological examination. Results of those assessments and medical records including imaging are reviewed at a consensus conference to identify cases of dementia and Alzheimer's disease using standardized research criteria. People who do not have dementia from this consensus process are also invited to enroll in the study. The final cognition scores from the CASI are derived using Item Response Theory (CASI-IRT), to improve score accuracy, measure cognitive change with less bias, and to account for missing test items (Crane et al., 2008; Ehlenbach et al., 2010; Li et al., 2017). Participants are prospectively followed until dementia incidence, drop-out, or death. Extensive health, lifestyle, biological, and demographic data are also collected.

As of March of 2020, the total ACT enrollment consisted of 5,763 participants. Our primary analysis was restricted 5,283 (92%) participants with a valid baseline CASI-IRT score, those who had lived in the exposure monitoring region (see below) at least 95% during the prior five years, and those with adjustment covariates (SI Figure S1 details participant retention). The analytic cohort had excellent residential histories and air pollution coverage (Blanco, Gassett, et al., 2022; Shaffer et al., 2021). On average, participants lived in the monitoring area >99% of the time, had exact geocoded residential addresses 98% of the time (e.g., vs. street level geocoding), and had imputed addresses only 5% of the time (i.e., from residential gaps).

Study procedures were approved by the University of Washington and Kaiser Permanente institutional review boards. ACT participants signed informed consent forms.

2.2 Air Pollution Exposure Assessment

We evaluated TRAP exposures for the ACT cohort using air pollution models from an extensive mobile monitoring campaign that was specifically designed for epidemiologic application (Blanco, Gassett, et al., 2022). In summary, the campaign monitored TRAP within a 1,200 land km^2 area in the greater Seattle area and consisted of repeated two-minute measurements at 309 roadside locations representative of ACT residential locations. Approximately 29 (IQR: 29-29, range: 26-35) measurements were collected from each location over the course of a year between March 2019 and March 2020 across all four seasons, all days of the week, and most hours of the day (5 AM – 11 PM). Median pollutant concentrations were estimated for each site visit, and these data were used to develop reference all-data site annual averages, as described below. This study focuses primarily on total PNC measures from the TSI NanoScan 3910, which measured 10-420 nm particles and is consistent with some air quality

recommendations to measure UFPs down to at least 10 nm (WHO, 2021). In sensitivity analyses, we investigate at 10-100 nm PNC from the NanoScan since UFPs are commonly defined as ≤ 100 nm and 20-1,000 nm PNC from the P-TRAK, a common UFP monitoring instrument. In secondary analyses, we evaluated nitrogen dioxide (NO_2) (Aerodyne Research Inc. CAPS NO_2).

As previously described (Blanco et al., 2023), we developed universal kriging – partial least squares (UK-PLS) models to predict PNC from hundreds of geographic covariates (e.g., land use, roadway proximity, population density). We evaluated the cross-validated model predictions against the all-data annual average estimates and summarized these with mean-square error (MSE) -based R^2 (R^2_{MSE}) in primary analyses and more traditional regression-based R^2 (R^2_{reg}) in secondary analyses. R^2_{MSE} evaluates the degree to which prediction-observation pairs are identical (i.e., along the one-to-one line) and is thus well suited to evaluate predictive performance, whereas R^2_{reg} evaluates linear associations. This UK-PLS model was slightly modified from previously developed models (Blanco et al., 2024; Blanco, Gassett, et al., 2022) to allow for comparisons across monitoring designs (described below), as detailed in Note S2.

We used this model to predict time-weighted average PNC exposures for each participant at baseline based on their prior five-year residential history. Details on the use of residential histories to evaluate exposures can be found in earlier work (Blanco et al., 2024; Shaffer et al., 2021). We assumed that UFP exposure surfaces had been stable over time since historical UFP data to develop longer-term exposures were limited. Our prior work shows that this assumption may be valid since the spatial variability of UFPs is much greater than the long-term temporal trend (2009+) and thus, capturing the spatial variability, as was done in the Seattle mobile monitoring campaign, is most critical (See Figure S8 in Blanco et al. [2024]). In sensitivity analyses, we restricted analyses to participants enrolled starting in 2010 to reduce to soften this assumption.

2.3 Restricted Air Pollution Mobile Monitoring Designs

To investigate the role of air pollution mobile monitoring design on health inferences, we subsampled the mobile monitoring data (309 roadside locations x ~29 visits each) with replacement following common, more restricted sampling designs, as described in

Table 1. We have previously described this general approach (Blanco et al., 2023), although this work included additional sampling designs, some of which were not repeated from earlier work.

Table 1.Reduced sampling designs implemented using mobile monitoring data from roadside sites (309 sites x ~29 visits each).[a]

Design[a]	Versions	No. of Versions	Total Visits	Visits per Site	Campaign Repetitions per Version
All-data	All-data	1	8,969	29[b]	1
Fewer Visits (no temporal restrictions)	4, 6, 12 visits per site	3	1,236, 1,854, 3,708	4, 6, 12	30
Fewer Seasons[c]	1-4 seasons	4	3,708	12	30
Fewer Hours	Weekday business or rush hours, unadjusted or temporally-adjusted	4	3,708	12	30
Unbalanced Visits	High (H) and low (L) variability sites receive the following visits: H2 L22, H6 L18, H12 L12 (all receive 12 visits), H18 L6, H22 L2	5	3,708	2-22 (avg: 12)	30

[a] The all-data design is a reference for all other designs, which have fewer site visits.
[b] mean and median: 29; IQR: 29-29; range: 26-35
[c] Samples were distributed evenly across the randomly selected seasons (e.g., 12 site visits/3 seasons = 4 site visits/season).

In the first design, we sampled each site a fewer number of times (n=4, 6, 12 visits) with no additional temporal restrictions, and conducted 30 "campaign" repetitions of each.

In the second design, we restricted sampling to fewer (1-3) seasons and collected 12 visits from each site, ensuring that samples were evenly distributed across sampling seasons.

In the third design, we sampled fewer visits per site (n=12) during weekday business (9 AM - 5 PM) or rush (7-10 AM and 3-6 PM) hours. The fewer visit design with 12 visits per site was a reference for these designs since it collected the same number of visits per site without temporal restrictions.

We used business and rush hour visit samples as is (unadjusted) or temporally-adjusted, a common approach for addressing known biases resulting from restricted sampling campaigns that do not sample during the full exposure period of interest (e.g., weekends, night time) (Eeftens et al., 2012; Klompmaker et al., 2015; Montagne et al., 2015; van de Beek et al., 2021; van Nunen et al., 2017). This approach generally entails using continuous air monitoring from a "background" or low-concentration location; calculating time-specific adjustment factors, based commonly on the difference between a time-specific (e.g., hourly) measurement and the site's long-term average; and applying these differences to the measured concentrations. Details of our approach are in Note S1. In summary, we: 1) simulated a long-term UFP monitoring site at an urban background site from collocated PNC and NO_2 measures along with temporal indicators since continuous UFP measures were otherwise unavailable for the entire

mobile monitoring study period; 2) generated adjustment factors, defined as the difference between the predicted hourly PNC and the long-term average PNC at that background site; and 3) applied these adjustment factors to the mobile monitoring data collected under business and rush hours designs.

In our fourth design, we simulated a practice employed by nearly all field campaigns where location samples are spatially unbalanced such that locations have different visit frequencies. Specifically, we sampled based on site variability, which was determined by regressing site-specific PNC interquartile ranges (IQR; median [range]: 7,183 [2,834-22,625] pt/cm^3 based on ~29 visits per site) against the first two PLS components, which summarized hundreds of geographic covariate predictors (see below for example covariates). The in-sample model R^2 was 0.46. We used this model to predict in-sample site-specific IQR and ordered these such that 129 (42%) sites were treated as medium variability sites where visits continued to be fixed to 12. The remaining sites were split into high (H) or low (L) variability (n=90 [29%] each). Figure S5 shows the distribution of IQRs used. We visited high-variability sites more times (14 to 22 visits) and low-variability sites fewer times (10 to 2 visits), and vice versa.

As with our all-data exposure models, we calculated annual average PNC site concentrations with the resulting data from each sampling campaign, developed exposure prediction models, and used these to assess participant exposures. In total, there were 480 exposure models for primary analyses using 10-420 nm PNC from the NanoScan. We compared each model's cross-validated site predictions against the all-data annual averages (our best estimates). We and others have shown the importance of validating model predictions against unbiased estimates since biased, unstable campaign measurements (e.g., from restricted sampling designs) can produce noisy and misleading conclusions (Blanco, Doubleday, et al., 2022; Blanco et al., 2023; Kerckhoffs et al., 2016; Messier et al., 2018).

We repeated this process for sensitivity and secondary analyses of other pollutants.

2.4 Inferential Analyses

We evaluated the association between cognitive function (CASI-IRT) and air pollution using the model:

$$Y_j = \alpha + \beta_m \hat{X}_{j,m} + \sum_i \delta_{m,i} W_{j,i} + \epsilon_j \quad (1)$$

where Y_j is the baseline CASI-IRT for participant j; α is the modeled intercept; $\hat{X}_{j,m}$ is the predicted five-year average air pollution exposure prior to baseline for participant j from exposure model m; β_m is the linear association between air pollution and CASI-IRT and is the main parameter of interest; and $W_{j,i}$ and $\delta_{m,i}$ are indexed by adjustment variable i with $\delta_{m,i}$ representing the respective estimated coefficient. In reduced models, adjustment variables included participant age, calendar year (2 yr categories), sex, and education (no degree, high school equivalent, bachelor's, master's, doctorate, other). Primary models further adjusted for race (White, People of Color) and socioeconomic status (SES) based on the Neighborhood Disadvantage Index (NDI) at a participant's longest-lived address at or prior to baseline. NDI is a validated indicator composed of census tract-level variables from the American Community

Survey (Miles et al., 2016). We contextualized these results by calculating age-equivalents for the primary analysis with the all-data reference exposure model, defined as the air pollution to age coefficient ratio. We used robust standard errors to improve the reliability of inference even if the model was misspecified (Zeileis, 2004; Zeileis et al., 2020; Zeileis & Hothorn, 2002). We used a p-value threshold of 0.05 to determine statistical significance.

All analyses were conducted in R (v. 4.2.2) (R Core Team, 2023).

3 Results

3.1 Cohort Characteristics

Table 2 describes the baseline analytic cohort characteristics. Mean age was 74 (SD: 6) years old, slightly more were female, about half had at least a college education, and the mean CASI-IRT score was 0.33 (SD: 0.71).

Table 2. Baseline cohort characteristics.[1]

	Low PNC (N=1744)	Medium PNC (N=1742)	High PNC (N=1797)	Overall (N=5283)
Visit Age (Years)				
Mean (SD)	73.6 (6.05)	74.1 (6.38)	74.5 (6.48)	74.1 (6.32)
Median [Min, Max]	72.0 [65.0, 98.0]	73.0 [65.0, 96.0]	73.0 [65.0, 101]	73.0 [65.0, 101]
Sex				
Male	749 (42.9%)	719 (41.3%)	730 (40.6%)	2198 (41.6%)
Female	995 (57.1%)	1023 (58.7%)	1067 (59.4%)	3085 (58.4%)
Degree				
None	126 (7.2%)	136 (7.8%)	155 (8.6%)	417 (7.9%)
GED/High School	648 (37.2%)	605 (34.7%)	728 (40.5%)	1981 (37.5%)
Bachelor's	414 (23.7%)	427 (24.5%)	395 (22.0%)	1236 (23.4%)
Master's	275 (15.8%)	308 (17.7%)	254 (14.1%)	837 (15.8%)
Doctorate	107 (6.1%)	114 (6.5%)	96 (5.3%)	317 (6.0%)
Other	174 (10.0%)	152 (8.7%)	169 (9.4%)	495 (9.4%)
CASI-IRT				
Mean (SD)	0.370 (0.688)	0.357 (0.720)	0.273 (0.711)	0.333 (0.707)
Median [Min, Max]	0.414 [-1.96, 1.75]	0.387 [-1.98, 1.75]	0.295 [-2.12, 1.75]	0.366 [-2.12, 1.75]
Residential PNC (pt/cm3) Exposure				
Mean (SD)	8750 (648)	10100 (310)	12500 (2080)	10500 (2010)

	Low PNC (N=1744)	Medium PNC (N=1742)	High PNC (N=1797)	Overall (N=5283)
Median [Min, Max]	8880 [5930, 9560]	10100 [9560, 10700]	11700 [10700, 22100]	10100 [5930, 22100]

[1]Low, medium, and high PNC tertile is based on the predicted PNC from the all-data exposure model.

3.2 Exposure Assessment and Model Performances

The median (IQR) site PNC (10-420 nm) from the all-data campaign was 9,747 (8,412-11,199) pt/cm^3. Figure S6 details this distribution, the concentration distributions of other PNC measures from sensitivity analyses which produced lower concentrations, and the estimated site concentrations from restricted mobile monitoring sampling campaigns.

The all-data PNC exposure model had a cross-validated $R^2_{MSE,Ref}$ of 0.65 (Figure S7). Reduced sampling designs produced incrementally worse predictions when campaigns had ≤12 visits per location (median $R^2_{MSE} \approx 80-90\%$ of $R^2_{MSE,Ref}$), unbalanced (variable) visit frequencies across locations ($R^2_{MSE} \approx 80-90\%$ of $R^2_{MSE,Ref}$), durations <1 year ($R^2_{MSE} \approx 70-90\%$ of $R^2_{MSE,Ref}$), and restricted sampling hours (M-F business, rush; $R^2_{MSE} \approx 60-90\%$ of $R^2_{MSE,Ref}$). Performance was generally worsened when temporal adjustments were applied to business and rush hour designs (see Figure S8 for R^2 comparisons and Figure S9 for site-specific prediction errors). Sensitivity and secondary analyses of 10-100 nm PNC, 20-1,000 nm PNC, and NO_2 as well as R^2_{reg} showed similar patterns (Figure S10-S11).

The median (IQR) predicted PNC participant exposure was 10,124 (9,293-11,100) pt/cm^3 and ranged from 5,930-22,134 pt/cm^3. Exposure predictions for sampling campaigns varied across campaigns (Figure S12). The business hour design tended to underpredict high exposures relative to the all-data exposure model, while the rush hour design overpredicted high exposures. Designs with few visits to high variability sites (H2 L22) had highly variable predictions across campaigns, particularly for high concentrations. Results showed similar patterns in sensitivity analyses of 10-100 nm and 20-1,000 nm PNC. Predictions from the all-data exposure model were strongly correlated with those from most sampling campaigns (median Pearson correlations [R] > 0.85), indicating similar exposure surface although the business hour design consistently predicted exposures with lower correlations (R ~0.77-0.78; Figure S13). All designs had atypical campaigns that had meaningfully lower correlations, indicating variability across campaign iterations and lower exposure model performances.

3.3 Inferential Analyses

In the reduced health inference model, the mean baseline CASI-IRT score decreased by -0.020 (95% CI: -0.036, -0.004) per 1,900 pt/cm^3 increment in PNC, equivalent to approximately 7.5 months of aging (Table S2). Health associations remained consistent across most reduced sampling designs including those with fewer visits (median point estimate: -0.19, IQR: -0.022, -0.016), fewer seasons (median: -0.19, IQR: -0.021, -0.016), and unbalanced sampling (median: -0.18, IQR: -0.022, -0.016) (Figure 2). However, unbalanced designs produced more variable

estimates across campaigns. Estimates from business and rush hour sampling designs were systematically attenuated compared to the reference. Median associations were -0.008 (IQR: -0.010, -0.006) for business hours and -0.012 (IQR: -0.014, -0.010) for rush hours. Temporally adjusting rush hour sampling designs brought the estimates closer to the reference, resulting in a median association of -0.015 (IQR: -0.017, -0.014), whereas adjustment did not meaningfully improve the business hours estimates. Figures S14 provides additional details, including the point estimates and 95% CIs of selected designs.

In the primary health inference model, the mean baseline CASI-IRT score increased by 0.002 (95% CI: -0.016, 0.020) per 1,900 pt/cm^3 increment in PNC, equivalent to approximately -0.7 months of aging (Table S2). As before, health associations from designs with fewer visits (median point estimate: 0.002, IQR: -0.001, 0.004), fewer seasons (median: 0.000, IQR: -0.001, 0.003), and unbalanced sampling (median: 0.001, IQR: -0.001, 0.004) produced similar estimates (Figure 2). Rush hour designs also yielded comparable results (median: 0.002, IQR: 0.000, 0.003), whereas business hour designs produced more positive estimates (median: 0.006, IQR: 0.005, 0.007). Temporal adjustment reduced these estimates, bringing the business hour estimates closer to the reference (median: 0.002, IQR: 0.001, 0.004) while pushing the rush hour estimates away from the reference (median: -0.003, IQR: -0.005, -0.001). We discuss these results further in the discussion in the context of exposure measurement error. Figure S15 shows the point estimates and 95% CIs of selected designs. Sensitivity analyses restricted to participants enrolled since 2010 produced similar results (Table S3).

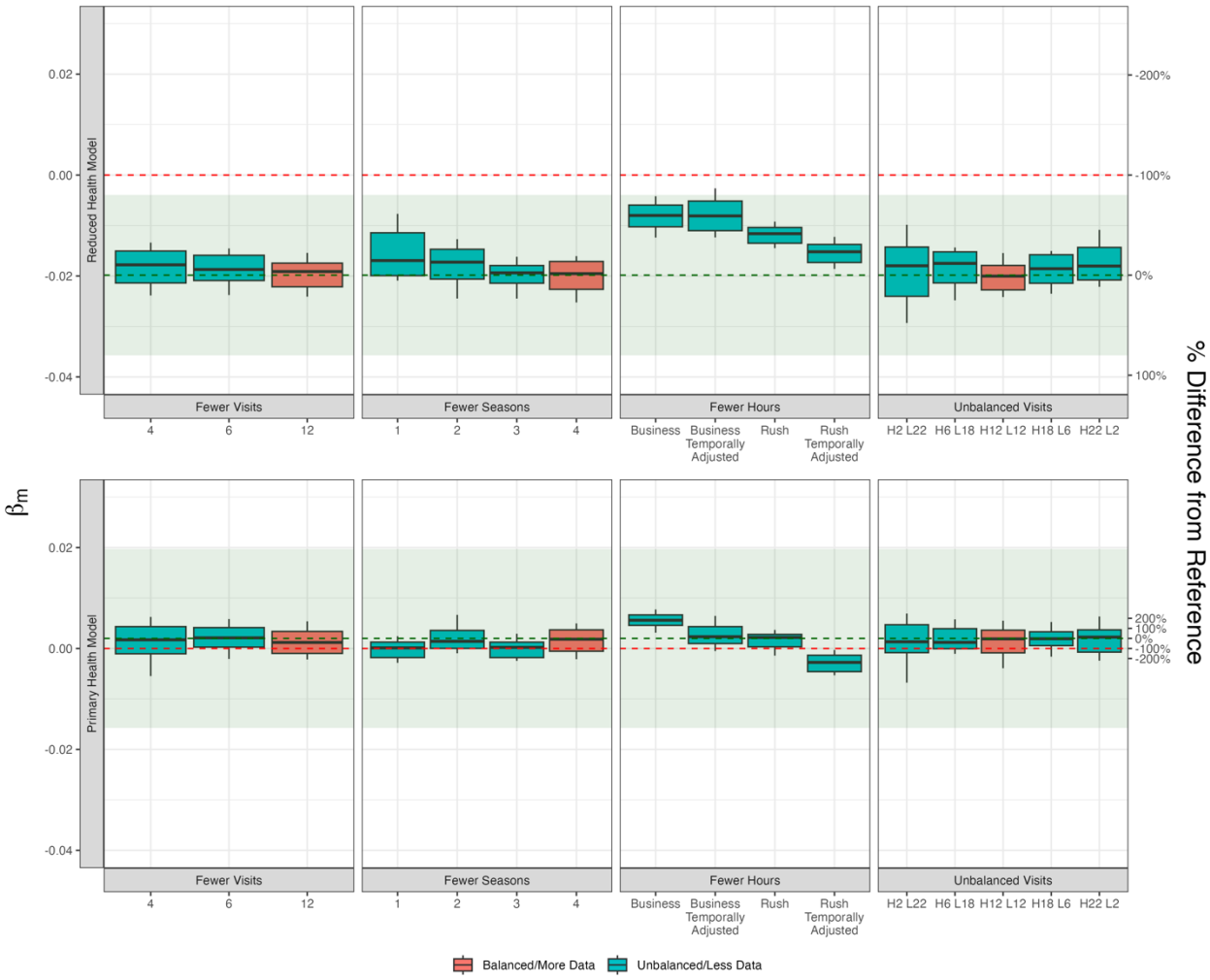

Figure 2. Estimated adjusted association between PNC (1,900 pt/cm^3) and cognitive function (CASI-IRT) from reduced (adjusted for age, calendar year, sex, education) and primary (models are further adjusted for race and SES) models. The green lines and areas indicate the health estimates from all-data exposure models, which are 0.002 (95% CI: -0.016, 0.020) in the primary model and -0.020 (95% CI: -0.036, -0.004) in the reduced model. Boxplots show the results when using exposure estimates from reduced mobile monitoring sampling campaigns (N=30 estimates per boxplot). Percents on the y-axis show the relative difference when compared to reference health estimates. Boxes show the median and IQR, whiskers show the 10th and 90th percentiles.

Figure S16 shows similar patterns for 10-100 nm and 20-1,000 nm PNC as well as NO_2.

4 Discussion

We found no strong evidence of an association between higher UFP exposures and lower late-life cognitive function in the ACT cohort. These findings are generally in line with the mixed results reported in the literature regarding the relationship between UFPs and brain health outcomes (Blanco et al., 2024; Brugge & Fuller, 2020; Gan et al., 2023; Weichenthal et al., 2020). Additionally, we observed that mobile monitoring campaigns can produce robust air pollution exposure models even with reduced sampling efforts, so long as data are collected

following spatially and temporally balanced approaches. Applying exposure surfaces from these reduced monitoring designs generally yielded health associations consistent with those from the all-data exposure model, with the exception of business and rush hour designs, which showed less consistent results.

A focus of this study was on evaluating the impact of air pollution mobile monitoring campaigns, which are increasingly implemented to address monitoring gaps (Kim et al., 2023), on health inferences. Guidance for epidemiologic applications has been limited (Blanco, Doubleday, et al., 2022; Blanco et al., 2023; Blanco, Gassett, et al., 2022; Doubleday et al., 2023), and monitoring approaches greatly vary. While the number of monitoring sites, visits, campaign duration, and sampling times can meaningfully impact exposure prediction models (Blanco, Doubleday, et al., 2022; Blanco et al., 2023), we further show how these differences translate to epidemiologic inferences.

When compared to the all-data monitoring campaign, campaigns with fewer visits (4-12) but no temporal restrictions, shorter durations (2-3 seasons), and a fixed number of visits across sites (12) produced only slightly worse exposure model performances (Figure S7); highly correlated (mostly >~0.85) participant exposures (Figure S13); and health inferences with only a small degree of variability (Figure 2). As expected, shorter campaigns (e.g., one season) and those with fewer repeat site visits (e.g., 4) generally had worse performing exposure models. Rush hour and especially business hour designs, had much worse exposure model performances, more variable participant exposure assessments, and health estimates that were inconsistent and generally more differed from the primary health effect estimates despite predicting exposures that were moderately (business) and highly (rush) correlated with exposures from the all-data model. Moreover, the health estimates associated with these monitoring campaigns were different from all other designs, including shorter (e.g., 1 season vs year-around) campaigns and those with fewer samples (e.g., 4 vs 12 site visits), suggesting that capturing temporal variability through extended hours designs might better summarize long-term annual average exposures. It's notable that these reduced day and hour campaigns are most common in the field since an operator is required to drive a vehicle and monitor instrumentation throughout the sampling period. Unfortunately, business hours designs may systematically capture lower-than-average air pollution times since they miss elevated rush hour and overnight concentrations. While rush hour designs collect data over fewer hours in a day, those hours may better represent longer-term averages in some situations.

Temporally adjusting business hours design data did not generally improve the health estimates (Figure 2, Figure S8, Figure S9). Prior work has also reported worse performing exposure models after applying a temporal adjustment when compared to not applying one at all (Chastko & Adams, 2019). This may be due to the heterogeneity in temporal patterns across different sites, which can be influenced by variable nearby sources such as airports, highways, or industrial areas (Blanco, Doubleday, et al., 2022). Applying a single site's temporal pattern to a larger study region may thus incorrectly or insufficiently adjust temporally restricted measures.

While there were some instances where health inferences were improved, the added complexity of temporal adjustment might have introduced classical-like measurement error,

reducing prediction accuracy. However, the improved temporal alignment may have decreased Berkson-like error, which was responsible for the dominant bias in health effect estimation from the unadjusted exposure estimates (Szpiro et al., 2011; Szpiro & Paciorek, 2013). In cases where inferences did not improve, the temporal adjustment might have introduced classical-like measurement error without sufficiently addressing or worsening Berkson-like error. Prior work has also shown that exposure measurement error can bias health inferences, and that the direction of the bias is unclear and not always towards the null (Bergen & Szpiro, 2015; Brenner & Loomis, 1994). Given that unbalanced monitoring designs may be challenging to improve with existing temporal adjustment approaches, these findings further highlight the importance of monitoring design, as reiterated below. Notably, the degree of bias will depend on true underlying exposure-outcome associations and realized monitoring design.

Interestingly, collecting less balanced visit data across sites generally resulted in worse and more variable exposure models, but this had a less pronounced impact on health inferences (Figure S7 and Figure 2). For epidemiologic applications, it may be acceptable to collect fewer visits at certain locations if logistical challenges arise, as long as the samples remain temporally balanced. It is important to note that there are various ways in which field sampling can lead to unbalanced data collection, and our approach reflects just one scenario based on the observed data. While fully balanced sampling designs are not always feasible in practice, they can be approximated in the field by using geographic information, among other methods. Defining target sites becomes increasingly complex when multiple pollutants are of interest since spatial patterns can vary greatly between different pollutants.

Spatial and temporal compatibility between monitoring and cohort locations (i.e., similarity in the spatial distribution and time period of interest) is an important component of minimizing the impact of measurement error on health inferences (Szpiro & Paciorek, 2013). Our study is inherently spatially aligned since the full mobile monitoring campaign was specifically designed to capture exposures for the ACT cohort (Blanco, Gassett, et al., 2022). Moreover, the design of the all-data campaign generally estimates unbiased annual average exposures, as we have previously shown (Blanco, Doubleday, et al., 2022). On the other hand, the day and time restricted designs sample during times that are temporally misaligned with the longer-term exposures of interest. These designs contribute to bias from Berkson-like error, which is the difference between the true annual average exposure surface and the more limited part captured by the modeling process (Szpiro & Paciorek, 2013). The ideal way to eliminate the bias from temporal misalignment is by modifying the sampling design. Future research can explore whether the available data collected can be leveraged to better support more comprehensive spatiotemporal modeling that better adjusts for and captures the underlying surface. While reweighting the data may be one simple approach, this may not be possible with very restricted designs like business hours approaches where large periods of critical time periods are missing.

More generally, our monitoring design findings are conservative with respect to realized campaigns, which may perform worse than the designs tested here. Field campaigns often incorporate multiple features that can limit their ability to assess long-term population exposures, including monitoring for less than a year, collecting few repeat visits per location

(median ~4), sampling exclusively during weekday business hours, and collecting a variable number of measurements per location (Kim et al., 2023).

This study was conducted in the ACT cohort, which has consistently and extensively collected cognitive outcomes, demographics, and other measures from older participants since its inception in 1994. More generally, these analyses are most relevant for older White populations who are fairly well educated given the demographics of the cohort. The low levels of air pollution in the region and the strong correlation between air pollution and SES (Hajat et al., 2015) may explain why we observed no strong evidence of an association between UFP and cognitive function in primary analyses. While there may be some residual confounding and we could have implemented more complex analyses, a focus of this study was on estimating the health inference bias and variability when using exposure models developed from common restricted sampling campaigns. Any analytic concerns would have remained constant across monitoring designs, thus allowing us to still capture trends. Moreover, our prior work investigating UFPs and dementia in ACT reported consistent results of no evidence of an association across an extensive set of sensitivity and secondary analyses, suggesting that our analytic choices had a limited impact on our results (Blanco et al., 2024).

A feature of our temporal adjustment approach is that it was based on a simulated UFP monitoring site, which produced good PNC predictions and captured much of the temporal variation observed for UFP (Figure S2 – Figure S4). Various temporal adjustment approaches are used in the literature, and these can themselves produce fluctuating adjustment factors. And while we used hourly adjustment factors to adjust two-minute mobile monitoring readings, we have previously shown that hourly and minute-level concentrations are highly correlated (Blanco, Doubleday, et al., 2022).

Notably, our use of 2019 UFP exposures inherently adds exposure measurement error to all our analyses, particularly for earlier time periods. Historical UFP data are limited, and we assumed that the exposure surface remained constant over time in these analyses (Blanco, 2021; Blanco et al., 2024; Kim et al., 2017; Levy et al., 2015; Meng et al., 2019; Molter et al., 2010; Wang et al., 2011). While our results were similar when we restricted analyses to more recent years, it will be important to replicate these and longitudinal analyses as longer-term UFP exposure models become available.

Finally, most mobile monitoring campaigns explicitly collect non-stationary, on-road data. These achieve higher spatial coverage than temporary roadside designs like the one used in this study, but they have unique challenges that should be investigated. This includes the collection of higher concentrations that are more influenced by high tailpipe emissions and at locations further from residences of interest; the collecting of fewer measurements per location (seconds vs minutes) that make for more unstable estimates; and initially poorer performing exposure models (Doubleday et al., 2023; Kim et al., 2023).

In conclusion, we saw no strong evidence of an association between higher UFP exposures and lower late-life cognitive function. Future studies should replicate this work as longer-term UFP exposure models become available. Thoughtful monitoring design can enhance exposure assessment and, while potentially less critical in some settings, can support epidemiologic analyses. Campaigns restricted to weekday business or rush hour designs produced worse exposures and less accurate health inferences. Campaigns might consider

prioritizing sampling beyond these times to support robust TRAP exposure assessments and subsequent health analyses.

5 Acknowledgements

Research described in this article was conducted under contract to the Health Effects Institute (HEI), an organization jointly funded by the United States Environmental Protection Agency (EPA) (Assistance Award No. CR-83998101) and certain motor vehicle and engine manufacturers. The contents of this article do not necessarily reflect the views of HEI, or its sponsors, nor do they necessarily reflect the views and policies of the EPA or motor vehicle and engine manufacturers. This research was also supported by NIA/NIEHS R01ES026187 to LS. We thank the ACT participants for the data they have provided and the many ACT investigators and staff who steward that data (NIA U19AG066567). You can learn more about ACT at: https://actagingstudy.org/. All statements in this report, including its findings and conclusions, are solely those of the authors and do not necessarily represent the views of the NIA or the NIH.

6 References

Apte, J. S., Messier, K. P., Gani, S., Brauer, M., Kirchstetter, T. W., Lunden, M. M., Marshall, J. D., Portier, C. J., Vermeulen, R. C. H., & Hamburg, S. P. (2017). High-Resolution Air Pollution Mapping with Google Street View Cars: Exploiting Big Data. *Environmental Science & Technology*, *51*, 6999–7008.

Austin, E., Xiang, J., Gould, T. R., Shirai, J. H., Yun, S., Yost, M. G., Larson, T. V., & Seto, E. (2021). Distinct Ultrafine Particle Profiles Associated with Aircraft and Roadway Traffic. *Environmental Science & Technology*, *55*(5), 2847–2858. https://doi.org/10.1021/acs.est.0c05933

Bergen, S., & Szpiro, A. A. (2015). Mitigating the impact of measurement error when using penalized regression to model exposure in two-stage air pollution epidemiology studies. *Environmental and Ecological Statistics*, *22*(3), 601–631. https://doi.org/10.1007/s10651-015-0314-y

Blanco, M. N. (2021). Traffic-Related Air Pollution and Dementia Incidence in a Seattle-Based, Prospective Cohort Study. Dissertation. *University of Washington*.

Blanco, M. N., Bi, J., Austin, E., Larson, T. V., Marshall, J. D., & Sheppard, L. (2023). Impact of Mobile Monitoring Network Design on Air Pollution Exposure Assessment Models. *Environmental Science & Technology*, *57*(1), 440–450. https://doi.org/10.1021/acs.est.2c05338

Blanco, M. N., Doubleday, A., Austin, E., Marshall, J. D., Seto, E., Larson, T. V., & Sheppard, L. (2022). Design and evaluation of short-term monitoring campaigns for long-term air

pollution exposure assessment. *Journal of Exposure Science & Environmental Epidemiology*, *33*(2), 465–473. https://doi.org/10.1038/s41370-022-00470-5

Blanco, M. N., Gassett, A., Gould, T., Doubleday, A., Slager, D. L., Austin, E., Seto, E., Larson, T. V., Marshall, J. D., & Sheppard, L. (2022). Characterization of Annual Average Traffic-Related Air Pollution Concentrations in the Greater Seattle Area from a Year-Long Mobile Monitoring Campaign. *Environmental Science & Technology*, *56*(16), 11460–11472. https://doi.org/10.1021/acs.est.2c01077

Blanco, M. N., Shaffer, R. M., Li, G., Adar, S. D., Carone, M., Szpiro, A. A., Kaufman, J. D., Larson, T. V., Hajat, A., Larson, E. B., Crane, P. K., & Sheppard, L. (2024). Traffic-related air pollution and dementia incidence in the Adult Changes in Thought Study. *Environment International*, *183*, 108418. https://doi.org/10.1016/j.envint.2024.108418

Brenner, H., & Loomis, D. (1994). Varied forms of bias due to nondifferential error in measuring exposure. *Epidemiology (Cambridge, Mass.)*, *5*(5), 510–517.

Brugge, D., & Fuller, C. H. (Eds.). (2020). *Ambient combustion ultrafine particles and health*. Nova Science Publishers.

Chastko, K., & Adams, M. (2019). Assessing the accuracy of long-term air pollution estimates produced with temporally adjusted short-term observations from unstructured sampling. *Journal of Environmental Management*, *240*(2019), 249–258. https://doi.org/10.1016/j.jenvman.2019.03.108

Crane, P. K., Narasimhalu, K., Gibbons, L. E., Mungas, D. M., Haneuse, S., Larson, E. B., Kuller, L., Hall, K., & van Belle, G. (2008). Item response theory facilitated cocalibrating cognitive

tests and reduced bias in estimated rates of decline. *Journal of Clinical Epidemiology*, *61*(10), 1018–1027. https://doi.org/10.1016/j.jclinepi.2007.11.011

Doubleday, A., Blanco, M. N., Austin, E., Marshall, J. D., Larson, T. V., & Sheppard, L. (2023). Characterizing Ultrafine Particle Mobile Monitoring Data for Epidemiology. *Environmental Science & Technology*, *57*(26), 9538–9547. https://doi.org/10.1021/acs.est.3c00800

Eeftens, M., Tsai, M. Y., Ampe, C., Anwander, B., Beelen, R., Bellander, T., Cesaroni, G., Cirach, M., Cyrys, J., de Hoogh, K., De Nazelle, A., de Vocht, F., Declercq, C., Dedele, A., Eriksen, K., Galassi, C., Gražulevičiene, R., Grivas, G., Heinrich, J., … Hoek, G. (2012). Spatial variation of PM2.5, PM10, PM2.5 absorbance and PMcoarse concentrations between and within 20 European study areas and the relationship with NO2—Results of the ESCAPE project. *Atmospheric Environment*, *62*(2012), 303–317. https://doi.org/10.1016/j.atmosenv.2012.08.038

Ehlenbach, W. J., Hough, C. L., Crane, P. K., Haneuse, S. J. P. A., Carson, S. S., Curtis, J. R., & Larson, E. B. (2010). Association between acute care and critical illness hospitalization and cognitive function in older adults. *JAMA*, *303*(8), 763–770. https://doi.org/10.1001/jama.2010.167

Flood-Garibay, J. A., Angulo-Molina, A., & Méndez-Rojas, M. Á. (2023). Particulate matter and ultrafine particles in urban air pollution and their effect on the nervous system. *Environmental Science. Processes & Impacts*, *25*(4), 704–726. https://doi.org/10.1039/d2em00276k

564 Gan, W., Manning, K. J., Cleary, E. G., Fortinsky, R. H., & Brugge, D. (2023). Exposure to ultrafine
565 particles and cognitive decline among older people in the United States. *Environmental*
566 *Research*, *227*, 115768. https://doi.org/10.1016/j.envres.2023.115768
567 Hajat, A., Hsia, C., & O'Neill, M. S. (2015). Socioeconomic Disparities and Air Pollution Exposure:
568 A Global Review. *Current Environmental Health Reports*, *2*(4), 440–450.
569 https://doi.org/10.1007/s40572-015-0069-5
570 HEI. (2013). *Understanding the Health Effects of Ambient Ultrafine Particles. Perspectives 3.*
571 https://www.healtheffects.org/publication/understanding-health-effects-ambient-
572 ultrafine-particles
573 Karumanchi, S., Siemiatycki, J., Richardson, L., Hatzopoulou, M., & Lequy, E. (2021). Spatial and
574 temporal variability of airborne ultrafine particles in the Greater Montreal area: Results
575 of monitoring campaigns in two seasons. *Science of The Total Environment*, *771*(2021),
576 144652. https://doi.org/10.1016/j.scitotenv.2020.144652
577 Kerckhoffs, J., Hoek, G., Messier, K. P., Brunekreef, B., Meliefste, K., Klompmaker, J. O., &
578 Vermeulen, R. (2016). Comparison of Ultrafine Particle and Black Carbon Concentration
579 Predictions from a Mobile and Short-Term Stationary Land-Use Regression Model.
580 *Environmental Science & Technology*, *50*(23), 12894–12902.
581 https://doi.org/10.1021/acs.est.6b03476
582 Kim, S.-Y., Blanco, M. N., Bi, J., Larson, T. V., & Sheppard, L. (2023). Exposure assessment for air
583 pollution epidemiology: A scoping review of emerging monitoring platforms and
584 designs. *Environmental Research*, *223*(2023).
585 https://doi.org/10.1016/j.envres.2023.115451

586 Kim, S.-Y., Olives, C., Sheppard, L., Sampson, P. D., Larson, T. V., Keller, J. P., & Kaufman, J. D.
587 (2017). Historical prediction modeling approach for estimating long-term concentrations
588 of PM2.5 in cohort studies before the 1999 implementation of widespread monitoring.
589 *Environmental Health Perspectives*, *125*(38–46). https://doi.org/10.1289/EHP131
590 Klompmaker, J. O., Montagne, D. R., Meliefste, K., Hoek, G., & Brunekreef, B. (2015). Spatial
591 variation of ultrafine particles and black carbon in two cities: Results from a short-term
592 measurement campaign. *Science of the Total Environment*, *508*(2015), 266–275.
593 https://doi.org/10.1016/j.scitotenv.2014.11.088
594 Knibbs, L. D., Cole-Hunter, T., & Morawska, L. (2011). A review of commuter exposure to
595 ultrafine particles and its health effects. *Atmospheric Environment*, *45*(16), 2611–2622.
596 https://doi.org/10.1016/j.atmosenv.2011.02.065
597 Kukull, W. A., Higdon, R., Bowen, J. D., McCormick, W. C., Teri, L., Schellenberg, G. D., Van Belle,
598 G., Jolley, L., & Larson, E. B. (2002). Dementia and Alzheimer disease incidence: A
599 prospective cohort study. *Archives of Neurology*, *59*(11), 1737–1746.
600 https://doi.org/10.1001/archneur.59.11.1737
601 Levy, I., Levin, N., Yuval, Y., Schwartz, J. D., & Kark, J. D. (2015). Back-extrapolating a land use
602 regression model for estimating past exposures to traffic-related air pollution.
603 *Environmental Science and Technology*, *49*(6), 3603–3610.
604 https://doi.org/10.1021/es505707e
605 Li, G., Larson, E. B., Shofer, J. B., Crane, P. K., Gibbons, L. E., McCormick, W., Bowen, J. D., &
606 Thompson, M. L. (2017). Cognitive Trajectory Changes Over 20 Years Before Dementia

607 Diagnosis: A Large Cohort Study. *Journal of the American Geriatrics Society*, *65*(12),
608 2627–2633. https://doi.org/10.1111/jgs.15077
609 Meng, J., Li, C., Martin, R. V., van Donkelaar, A., Hystad, P., & Brauer, M. (2019). Estimated
610 Long-Term (1981–2016) Concentrations of Ambient Fine Particulate Matter across
611 North America from Chemical Transport Modeling, Satellite Remote Sensing, and
612 Ground-Based Measurements. *Environmental Science & Technology*, *53*(9), 5071–5079.
613 https://doi.org/10.1021/acs.est.8b06875
614 Messier, K. P., Chambliss, S. E., Gani, S., Alvarez, R., Brauer, M., Choi, J. J., Hamburg, S. P.,
615 Kerckhoffs, J., Lafranchi, B., Lunden, M. M., Marshall, J. D., Portier, C. J., Roy, A., Szpiro,
616 A. A., Vermeulen, R. C. H., & Apte, J. S. (2018). Mapping Air Pollution with Google Street
617 View Cars: Efficient Approaches with Mobile Monitoring and Land Use Regression.
618 *Environmental Science and Technology*, *52*(21), 12563–12572.
619 https://doi.org/10.1021/acs.est.8b03395
620 Miles, J. N., Weden, M. M., Lavery, D., Escarce, J. J., Cagney, K. A., & Shih, R. A. (2016).
621 Constructing a Time-Invariant Measure of the Socio-economic Status of U.S. Census
622 Tracts. *Journal of Urban Health*, *93*(1), 213–232. https://doi.org/10.1007/s11524-015-
623 9959-y
624 Molter, A., Lindley, S., de Vocht, F., Simpson, A., & Agius, R. (2010). Modelling air pollution for
625 epidemiologic research – Part II: Predicting temporal variation through land use
626 regression. *Science of The Total Environment*, *409*(1), 211–217.
627 https://doi.org/10.1016/j.scitotenv.2010.10.005

628 Montagne, D. R., Hoek, G., Klompmaker, J. O., Wang, M., Meliefste, K., & Brunekreef, B. (2015).
629 Land Use Regression Models for Ultrafine Particles and Black Carbon Based on Short-
630 Term Monitoring Predict Past Spatial Variation. *Environmental Science and Technology*,
631 *49*(14), 8712–8720. https://doi.org/10.1021/es505791g
632 R Core Team. (2023). *R-A Language and Environment for Statistical Computing*. https://www.R-
633 project.org
634 Shaffer, R. M., Blanco, M. N., Li, G., Adar, S. D., Carone, M., Szpiro, A. A., Kaufman, J. D., Larson,
635 T. V., Larson, E. B., Crane, P. K., & Sheppard, L. (2021). Fine Particulate Matter and
636 Dementia Incidence in the Adult Changes in Thought Study. *Environmental Health*
637 *Perspectives*, *129*(8), 087001. https://doi.org/10.1289/EHP9018
638 Sheppard, L., & Blanco. (Under Review). *Optimizing Exposure Assessment for Inference about*
639 *Air Pollution Effects with Application to Cognitive Function*.
640 Szpiro, A. A., & Paciorek, C. J. (2013). Measurement error in two-stage analyses, with
641 application to air pollution epidemiology. *Environmetrics*, *24*(8), 501–517.
642 https://doi.org/10.1002/env.2233
643 Szpiro, A. A., Sheppard, L., & Lumley, T. (2011). Efficient measurement error correction with
644 spatially misaligned data. *Biostatistics*, *12*(4), 610–623.
645 https://doi.org/10.1093/biostatistics/kxq083
646 Teng, E. L., Hasegawa, K., Homma, A., Imai, Y., Larson, E., Graves, A., Sugimoto, K., Yamaguchi,
647 T., Sasaki, H., Chiu, D., & White, L. R. (2004). The Cognitive Abilities Screening
648 Instrument (CASI): A Practical Test for Cross-Cultural Epidemiological Studies of

649 Dementia. *International Psychogeriatrics*, *6*(1), 45–58.
650 https://doi.org/10.1017/s1041610294001602
651 US EPA. (2019). *Integrated science assessment (ISA) for particulate matter*. US Environmental
652 Protection Agency Washington, DC.
653 https://cfpub.epa.gov/ncea/isa/recordisplay.cfm?deid=347534
654 van de Beek, E., Kerckhoffs, J., Hoek, G., Sterk, G., Meliefste, K., Gehring, U., & Vermeulen, R.
655 (2021). Spatial and Spatiotemporal Variability of Regional Background Ultrafine Particle
656 Concentrations in the Netherlands. *Environmental Science & Technology*, *55*(2), 1067–
657 1075. https://doi.org/10.1021/acs.est.0c06806
658 van Nunen, E., Vermeulen, R., Tsai, M.-Y., Probst-Hensch, N., Ineichen, A., Davey, M., Imboden,
659 M., Ducret-Stich, R., Naccarati, A., & Raffaele, D. (2017). Land use regression models for
660 ultrafine particles in six European areas. *Environmental Science & Technology*, *51*(6),
661 3336–3345.
662 Wang, Y., Hopke, P. K., Chalupa, D. C., & Utell, M. J. (2011). Long-term study of urban ultrafine
663 particles and other pollutants. *Atmospheric Environment*, *45*(40), 7672–7680.
664 https://doi.org/10.1016/j.atmosenv.2010.08.022
665 Weichenthal, S., Olaniyan, T., Christidis, T., Lavigne, E., Hatzopoulou, M., Van Ryswyk, K.,
666 Tjepkema, M., & Burnett, R. (2020). Within-city Spatial Variations in Ambient Ultrafine
667 Particle Concentrations and Incident Brain Tumors in Adults. *Epidemiology*, *31*(2), 177.
668 https://doi.org/10.1097/EDE.0000000000001137
669 Weichenthal, S., Van Ryswyk, K., Goldstein, A., Shekarrizfard, M., & Hatzopoulou, M. (2016).
670 Characterizing the spatial distribution of ambient ultrafine particles in Toronto, Canada:

A land use regression model. *Environmental Pollution*, *208*(Pt A), 241–248. https://doi.org/10.1016/j.envpol.2015.04.011

WHO. (2021). *World Health Organization global air quality guidelines: Particulate matter (PM2.5 and PM10), ozone, nitrogen dioxide, sulfur dioxide and carbon monoxide*. World Health Organization. https://apps.who.int/iris/handle/10665/345329

Zeileis, A. (2004). Econometric Computing with HC and HAC Covariance Matrix Estimators. *Journal of Statistical Software*, *11*(10). https://doi.org/10.18637/jss.v011.i10

Zeileis, A., & Hothorn, T. (2002). Diagnostic Checking in Regression Relationships. *R News*, *2*(3), 7–10.

Zeileis, A., Köll, S., & Graham, N. (2020). Various Versatile Variances: An Object-Oriented Implementation of Clustered Covariances in *R*. *Journal of Statistical Software*, *95*(1). https://doi.org/10.18637/jss.v095.i01