

ARTICLE OPEN



Influence of on-road mobile monitoring design on ultrafine particle exposure models and cognitive health inferences

Magali N. Blanco¹✉, Annie Doubleday¹, Adam A. Szpiro², Julian D. Marshall³, Paul K. Crane⁴ and Lianne Sheppard^{1,2}

© The Author(s) 2026

BACKGROUND: On-road mobile monitoring is increasingly used to assess air pollution, but the implications of monitoring and analytic decisions on exposure prediction and health inferences remain unclear.

OBJECTIVE: This study evaluated the influence of on-road monitoring approaches in environmental epidemiology, specifically ultrafine particle (UFP) exposures and late-life cognitive function.

METHODS: We used data from a Seattle-based mobile monitoring campaign to develop a reference roadside UFP exposure model based on repeated measurements at 309 roadside locations and examine associations with cognitive function (Cognitive Abilities Screening Instrument—Item Response Theory [CASI-IRT]), in the Adult Changes in Thought cohort ($N = 5283$). To evaluate alternative designs, we subsampled on-road UFP measurements collected along 600 km of roadways, varying location visit frequencies, spatial balancing, and sampling times. New UFP models, some incorporating temporal and plume adjustments, were developed using universal kriging with partial least squares and used to estimate associations between UFP and CASI-IRT, after adjusting for age, year, sex, education, race, and socioeconomic status.

RESULTS: Using the reference exposure model in the primary health model, the mean baseline CASI-IRT score increased by 0.007 (95% CI: $-0.013, 0.027$) per 1900 pt/cm^3 increment in PNC. Associations were similar but relatively attenuated for all on-road sampling designs. Route-based sampling (which accounted for logistical field constraints and spatiotemporal correlation in the data) and very short (4- vs 12-visit) campaigns produced more variable health estimates. Applying temporal and plume adjustments had a minimal impact on the inferential results.

SIGNIFICANCE: In analyses where no association was observed between UFP and cognitive function, the on-road monitoring design produced similar but slightly attenuated point estimates. Secondary analyses with a reduced health model, which indicated a statistically significant but potentially confounded association, suggest that on-road design—particularly monitoring beyond weekday business hours—may have greater implications in other contexts.

IMPACT STATEMENT: Mobile monitoring is increasingly used to develop air pollution exposure models, yet the influence of monitoring design on health inferences remains unclear. Using extensive ultrafine particle (UFP) data from a monitoring campaign and health measures from a long-standing cohort study, we assess how on-road campaigns can be designed for epidemiologic research. We evaluate the effects of visit number, spatial balance, time selection, temporal adjustment methods, and plume adjustments on exposure and health models, providing guidance for mobile monitoring design in air pollution health research.

Keywords: Air pollution; Study designs; Exposure assessment; Epidemiology

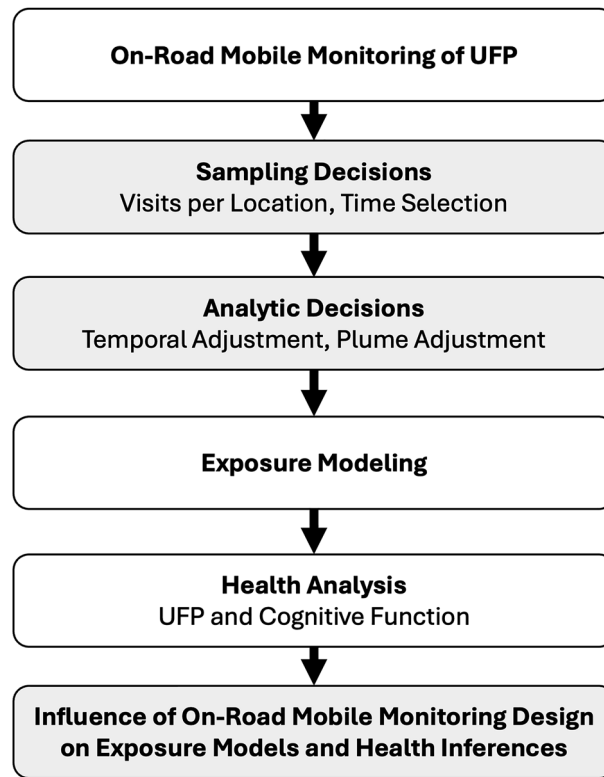
Journal of Exposure Science & Environmental Epidemiology (2026) 36:575–584; <https://doi.org/10.1038/s41370-026-00845-y>

¹Department of Environmental and Occupational Health Sciences, University of Washington, Seattle, WA, USA. ²Department of Biostatistics, University of Washington, Seattle, WA, USA. ³Department of Civil & Environmental Engineering, University of Washington, Seattle, WA, USA. ⁴Department of Medicine, University of Washington, Seattle, WA, USA. ✉email: magali@uw.edu

Received: 25 February 2025 Revised: 20 January 2026 Accepted: 17 February 2026
Published online: 7 March 2026

Graphical Abstract

Tweetable Abstract: How does mobile monitoring design impact health studies? We assess UFP data and health measures to evaluate visit number, spatial balance, timing, and adjustments to assess their influence on exposure and health models.



INTRODUCTION

Accurate air pollution exposure assessment is critical for environmental epidemiology. Mobile monitoring, the collection of repeated short-term samples from areas of interest, can be used to characterize poorly measured pollutants with high spatial resolution [1]. Traffic-related ultrafine particles (UFPs, ≤ 100 nm) are of growing interest given the accumulating evidence linking these to adverse cardiovascular, respiratory, and cognitive outcomes [2–5]. Still, the key design principles necessary to optimize on-road mobile monitoring for both exposure assessment and epidemiology remain unclear.

Several studies have evaluated how the number of repeat samples (site visits) impacts exposure estimates. Results may vary by pollutant, location, and averaging period, but generally, more measurements per location produce more precise concentration estimates. For example, in Oakland, CA, Messier et al. suggested that 4–8 visits per location for black carbon (BC) or nitrogen monoxide (NO) were sufficient [6], whereas Kerckhoffs et al. reported a number closer to 5–15 visits, depending on the modeling approach [7]. In Montreal, Canada, Hatzopoulou et al. found that 10–12 visits were necessary for UFP and nitrogen dioxide (NO₂) [8]. Similarly, in Bengaluru, India, Upadhyay et al. recommend 10 visits for UFP and BC [9].

Numerous gaps remain. We have previously demonstrated using fixed-site and stationary roadside monitoring data that limiting monitoring to weekday daytime hours leads to poorer exposure estimates and health inferences [10–12]. This has not been replicated for on-road monitoring campaigns, however, which measure pollutants at a greater number of unique locations and over shorter periods of time at any (less precise) location. Moreover, air pollution monitoring campaigns often lack spatially balanced measurements due to logistical constraints or intentional design choices, leading to variable visit frequencies. Additionally, while previous sampling studies have analyzed data

at the segment level, allowing for independent sampling of neighboring segments, real-world campaigns often measure neighboring road segments consecutively along a route, resulting in the measurements for neighboring segments being temporally correlated.

There are also important analytic considerations for on-road campaigns that have not been evaluated. For instance, many campaigns apply temporal adjustments to temporally imbalanced data (e.g., from weekday business hours) to make the data more closely resemble longer-term averages of interest. A few studies have evaluated the effectiveness of these approaches using stationary or roadside monitoring data, finding that these can sometimes worsen exposure estimates [12, 13] and may not substantially improve health inferences [12]. Replicating this work for on-road campaigns, which offer limited temporal coverage at any given location, is critical.

Moreover, many studies [14–17], though not all [18], have reported higher air pollution concentrations from on-road campaigns when compared to fixed-site locations. On-road campaigns that measure in traffic or near to other vehicles have a higher likelihood of unintentionally over-sampling from the high concentrations found in exhaust plumes. Some investigations explicitly focus on on-road vehicle plumes, using on-road monitoring campaigns to follow specific vehicles and/or measure vehicle emission factors [9, 19–24]. Epidemiologic studies, however, are often more concerned with residential exposures, where on-road plume concentrations are generally less relevant. To decrease plume concentrations, methods such as plume detection [15], the use of stationary measurements [25], and the application of multi-pollutant measurements [14] have been developed. Most studies, however, have not implemented these approaches, and it remains unclear whether they necessarily improve exposure estimates and health inferences.

We leverage an extensive on-road and roadside mobile monitoring campaign [26] along with late-life cognitive function measures from the Adult Changes in Thought (ACT) cohort [27] to address key gaps in on-road mobile monitoring for environmental epidemiology. The primary objectives of this study were to evaluate the impact of on-road monitoring designs and analytic approaches on subsequent (1) exposure estimates and (2) epidemiologic inferences, specifically associations between UFP exposures and cognitive function. This study builds on prior work focused on stationary roadside monitoring but addresses distinct challenges specific to on-road mobile campaigns, including several that are not applicable to stationary designs. These include the use of route-based data collection, the elevated risk of spatial and temporal imbalance given the larger number of monitoring locations involved, and strategies to address potential limitations—such as temporal and plume adjustments—that are unique to on-road monitoring. These distinctions motivate a separate evaluation and provide a complementary contribution by highlighting key design principles for improving exposure assessment models and health inferences from on-road campaigns.

METHODS

Study design

We evaluated the influence of on-road monitoring design on UFP exposure models and subsequent associations with cognitive function in the ACT cohort. This study consisted of the following, as summarized in Fig. 1:

- Sampling on-road UFP measures following common monitoring approaches from an extensive monitoring campaign.
- Applying adjustments to account for temporally imbalanced and plume-impacted measurements.
- Developing UFP exposure prediction models with the sampled data.
- Evaluating the association between predicted UFP exposures and cognitive function in ACT.
- Comparing exposure and health estimates to those from a reference stationary roadside monitoring campaign.

The ACT cohort and outcome ascertainment

We used cognitive function measures from the ACT cohort, as previously described [12]. In summary, ACT is a community-based, prospective cohort study investigating the aging brain in the greater Seattle area since 1994 [27]. The study randomly invites elderly individuals (65+) from Kaiser Permanente Washington to participate, assessing cognitive function at baseline using the Cognitive Abilities Screening Instrument (CASI), which evaluates attention, memory, language, and other cognitive functions [28]. Participants with high scores ($\geq 86/100$) are enrolled, while those with lower scores undergo further neuropsychological testing. Final CASI scores are derived using Item Response Theory (CASI-IRT) to enhance accuracy and account for missing items [29–31]. As of March 2020, the ACT study enrolled 5763 participants. This analysis included 5283 (92%) participants enrolled between 1994 and 2020, primarily excluding those who resided outside of the exposure monitoring region. Specifically, participants were included if they had a valid CASI-IRT score (5696, 99%), lived in the exposure monitoring region for at least 95% of the prior five years (5409, 94%), and had complete covariate information for adjustment (5283, 92%). Figure S1 in Blanco et al. 2025 further details participant retention.

Mobile monitoring campaign

We analyzed UFP measurements from a previously described Seattle mobile monitoring campaign [14, 26]. Monitoring was conducted both on-road and at temporary (2-min) roadside stationary locations within a 1200 km² area. Sampling occurred between March 2019 and March 2020, covering all four seasons, all days of the week, and most hours of the day (5 AM–11 PM). The overall campaign included repeat samples from approximately 600 unique roadway kilometers along nine fixed routes and 309 roadside locations representative of the cohort (see below for exclusions applied). Monitoring locations were selected to maximize spatial compatibility with the cohort. Specifically, the monitoring region was defined to include areas where most cohort members had historically

resided. Individual roadside sites were then selected to minimize the distance between monitored and cohort locations while ensuring that the spatial distribution of monitoring sites reflected that of the cohort. Additional details are provided in Blanco, Gasset et al. [26]. This study focuses on total UFP (particle number concentration [PNC]) using the TSI P-TRAK 8525 instrument (size range: 20–1000 nm), which operated at a high measurement frequency of 1 Hz, enabling accurate PNC measurements during on-road monitoring. In contrast, prior work using stationary roadside measurements has employed the TSI NanoScan 3910, which captured smaller particles of interest (size range: 10–420 nm) but with a slower response rate of approximately 0.017 Hz (one measurement every 60 s), making it less suitable for on-road applications [12].

We divided the driving routes into 100 m segments and assigned all on-road measurements to the segment midpoint. We excluded A1 roads (interstates and highways with restricted access) to better represent residential exposures; segments with fewer than a median of five 1-s measurements per visit; segments with fewer than 23 repeat visits; and segments immediately before or after a roadside location, as these were not fully on-road measures. The 1-Hz PNC measurements were averaged over 10-s periods to align with other instruments, and the median PNC was calculated across all 10-s measures within each visit to a segment. Then for each segment, we winsorized these median values across visits: values below the 2.5th and above the 97.5th percentiles were set to those thresholds, to reduce the influence of extreme values. Overall, the median number of visits per road segment ($N = 5874$) was 28 (IQR: 27–28) [14].

We then adjusted the on-road data for high concentrations (“plumes”) using a previously described approach [14]. This method applies absolute principal component score (APCS) to identify on-road source components based on pollutants measured alongside UFP (BC, NO₂, carbon dioxide [CO₂], and other UFP measures). The contributions from these factors were then reduced to estimate adjusted UFP concentrations, with the final adjustment selected through an iterative process that leveraged data less likely to be plume-impacted, specifically roadside measures from the same campaign.

We used the resulting data to calculate the mean visit concentrations of each road segment to represent the annual average exposure estimate at that location.

Measures from roadside locations were processed similarly, with each of the 309 sites having approximately 28 (IQR: 28–28) winsorized median visit concentrations from the P-TRAK instrument [10]. These data were used to develop site-specific annual averages, which were treated as external reference measures for the on-road prediction models, as described below.

Restricted air pollution mobile monitoring designs

We conducted hypothetical sampling campaigns to demonstrate how on-road mobile monitoring and analytic decisions influence exposure and health models. These included opportunistic designs with a limited number of visits per location and sampling restricted to typical working hours. We also evaluated the spatial imbalance that can arise when some locations receive more repeated measurements than others. Finally, we assessed whether known temporal imbalances and the presence of high-concentration plume events—potentially less representative of residential exposures—can be addressed analytically to improve the utility of on-road monitoring data for exposure assessment and health inference. Specifically, we sampled segment-level median UFP concentrations from drive-passes with replacement in the following process:

- **Visits per location:**
 - *4 visits:* We collected 4 visits per road segment location.
 - *12 visits:* We collected 12 visits per road segment location.
- **Spatially balanced sampling:** Visits were collected in spatially balanced or unbalanced approaches. Balanced approaches collected the same number of visits at all locations. Unbalanced approaches collected visit counts at each location that were guided by a lognormal distribution (e.g., mean: log 4, SD: log 2, range: 1–28) whereby some locations received more visits than others. We further considered spatial groupings by sampling at the route level, sampling at the segment level without grouping, sampling spatial clusters, or sampling based on road type, as described below.
 - *Spatially balanced routes:* All segments along a route were visited the same number of times and on the same day, thereby accounting for the temporal correlation inherent in field campaigns.

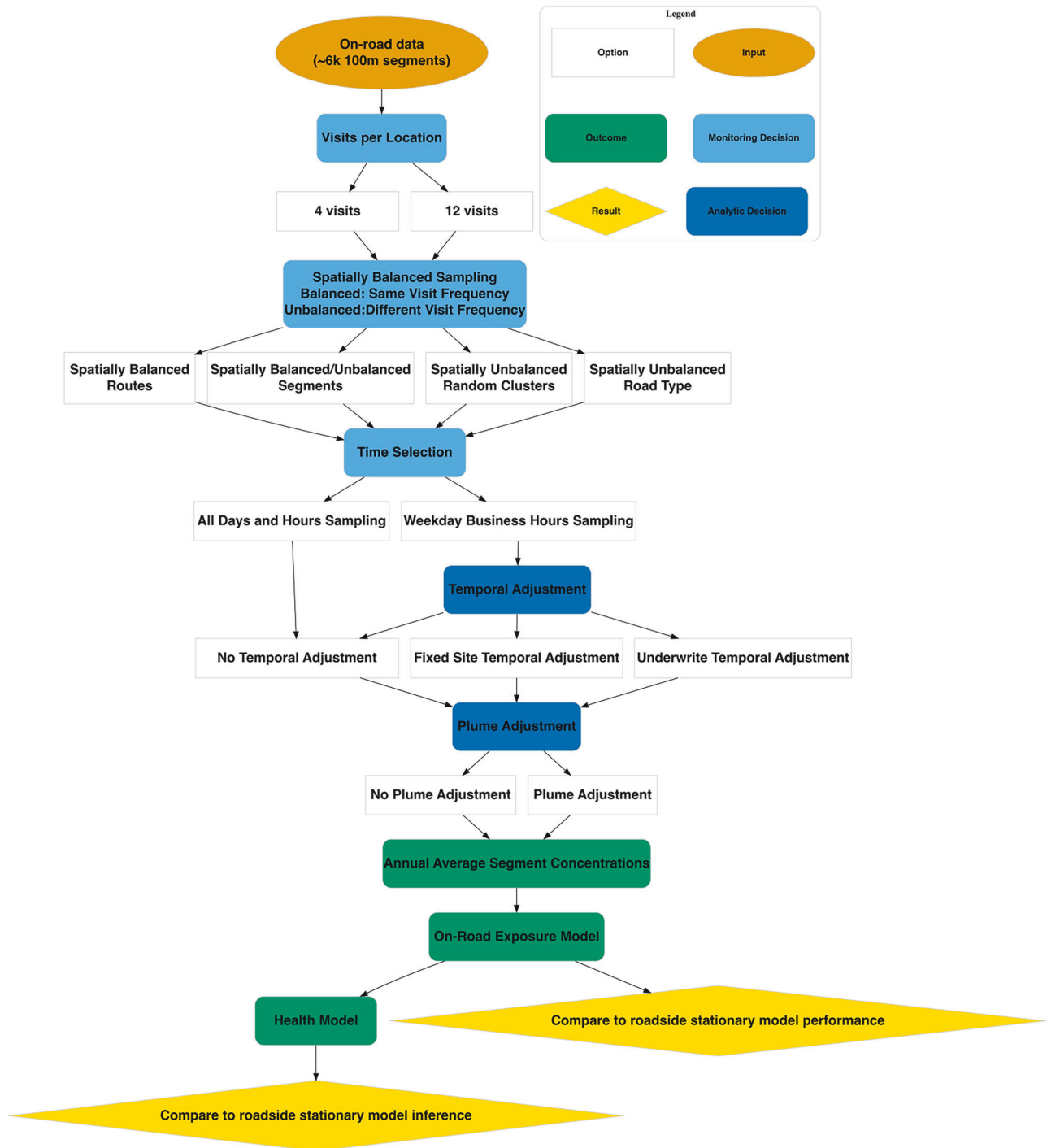


Fig. 1 Flow diagram illustrating the process of evaluating on-road mobile monitoring design for epidemiology. The process includes monitoring decisions (such as the number of visits per location, spatial balance related to how locations are sampled, and sampling times), analytic decisions (including temporal and plume adjustments), exposure model development, and health analyses. We evaluate the robustness of exposure and health outcomes by comparing results to those from a roadside reference campaign. There are a total of 112 design combinations; each is repeated 30 times (campaigns) for a total of 3360 sampling campaigns in the main analyses.

- *Spatially balanced segments*: Each road segment was visited the same number of times, but we allowed for independent segment-level visits (e.g., on different days, not based on route), a common evaluation approach, but reflecting a design not logistically possible (i.e., the vehicle would have to jump around randomly in space).
- *Spatially unbalanced segments*: Same as *spatially balanced segments*, but we allowed some segments to be visited more frequently than others.
- *Spatially unbalanced random clusters*: Some segments were sampled more heavily by establishing spatial clusters, each composed of an average of 93 100-m segments (see SI Fig. S1). Segments within a cluster were sampled the same number of times and allowed to be independent (i.e., not based on route).
- *Spatially unbalanced road type*: Road segments were sampled based on the predominant road type, with larger, more heavily trafficked roads receiving more visits.

- **Time selection:**

- *All-day and hour sampling:* Visits were conducted during all days and hours.
- *Weekday business hours sampling:* Visits were restricted to weekday business hours (9 AM–5 PM). For designs that sampled at the route (vs segment) level, more than 60% of the route was required to have been sampled during these hours.

- **Temporal adjustment:**

- *No temporal adjustment:* No temporal adjustment was applied to the resulting data.
- *Fixed-site temporal adjustment:* We followed a common approach of using a fixed-site background monitor to develop time-specific adjustment factors [32–35]. This method involves estimating time-specific adjustment factors based on PNC deviations at any given time (e.g., one hour in the study period) from the long-term average concentration at a background site. Since continuous PNC P-TRAK measures were not available for the entire study period, we simulated these from a fitted model of collocated P-TRAK PNC and NO₂ measures. We applied the resulting adjustment factors to the temporally unbalanced data. This approach assumes that temporal fluctuations in PNC concentrations at the background site reflect the temporal fluctuations of each monitored site, and that the adjusted concentrations better represent the long-term exposure of interest for epidemiologic inference. Additional methodological details are presented in Blanco et al. 2025 and the corresponding supplemental material.
- *Underwrite temporal adjustment:* We simulated a low concentration “background” site using the measures collected during a given campaign following an approach adapted from Hankey & Marshall [36, 37]. Specifically, for each campaign, we used the collected 1-s PNC data to estimate background levels by calculating the first percentile of three-hour rolling averages. We then computed both hourly and long-term (~1 year) average background PNC values. Hourly temporal adjustments were estimated as the difference between hourly and long-term background PNC, and these adjustments were applied to the mobile monitoring data. See Note S1 for additional details on this approach.

- **Plume adjustment:**

- *No plume adjustment:* We used the raw measures.
- *Plume adjustment:* We applied the plume adjustment detailed by Doubleday et al. [11] and summarized in the Methods. This method reduces the influence of transient, high on-road concentrations and produces adjusted concentrations that may be more representative of longer-term ambient conditions under the assumption that unusual spikes in pollutant concentrations may not reflect typical residential exposures.

In sensitivity analyses, we visited routes more times (20 visits) and used larger spatial clusters (Fig. S1).

For each campaign, we calculated the annual average concentration for each sampled road segment. We repeated this process 30 times to increase robustness, resulting in a total of 3360 sampling campaigns (112 designs × 30 campaigns each) in the main analyses. The annual averages from each campaign were used to develop an exposure prediction model, which was subsequently applied to evaluate the association between UFPs and cognitive function, as described below.

Air pollution exposure assessment

We developed exposure prediction models for the annual average roadside PNC reference model ($n = 309$), as previously described [12], as well as from each on-road sampling campaign. Models for PNC were developed using universal kriging—partial least squares (UK-PLS) and hundreds of geographic covariates predictive of TRAP (e.g., land use, roadway proximity, population density) [10].

We evaluated each on-road exposure model by predicting PNC at out-of-sample stationary roadside sites and comparing these predictions against the observed annual averages for those locations. We employed this approach because we hypothesized that the short-term stationary roadside measurements were more representative of residential exposures (thus more suited for epidemiologic application) than the on-road

measures because: (1) they were physically further from the center of the road and closer to residential facades; (2) they may be less impacted by on-road traffic plumes; (3) the measurement duration per location is slightly longer than for on-road locations (e.g., minutes vs. seconds), thus potentially resulting in more stable long-term averages; and (4) measures were collected and linked to exact (stationary) locations rather than the midpoints of longer 100 m road segments. We evaluated model performance using mean-square error (MSE)-based R^2 (R^2_{MSE}), which evaluates whether pairs of observations and predictions are identical, as well as using more common regression-based R^2 (R^2_{reg}), which evaluates whether pairs are linearly associated.

We used each campaign exposure model to predict time-weighted average PNC exposures for each ACT participant at baseline based on their prior five-year residential history. Given the limited availability of PNC data, this analysis assumed that the 2019–2020 PNC exposure surfaces reflected longer-term exposures and remained stable over time [38–44] and were therefore applicable to the cohort. Our prior work evaluating air pollution and cognitive function in the ACT cohort yielded similar results when analyses were restricted to more recent enrollment periods closer to 2019 [12]. This is further addressed in the Discussion.

Inferential analyses

We used data from ACT (described in 2.2 The ACT Cohort and Outcome Ascertainment) to evaluate the association between baseline cognitive function (CASI-IRT) and predicted PNC exposure from each exposure model using the following:

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

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 estimated association between air pollution and CASI-IRT and 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. Reduced models adjusted for participant age, calendar year, sex, and education. Primary models further adjusted for race and socioeconomic status based on the Neighborhood Disadvantage Index [45] of each participant's longest-lived address at or prior to baseline. We used robust standard errors (sandwich estimator) to improve the reliability of our inferences [46–48] and a p-value threshold of 0.05 to determine statistical significance.

We compared the estimated health association ($\hat{\beta}_m$) from applying each on-road exposure model to the association estimated from using the reference roadside exposure model.

We conducted all analyses in R (v. 4.2.2) [49].

RESULTS

Cohort characteristics

ACT participants had an average age of 74 years (SD: 6), with a slight majority (58%) being female. Over half of the participants held a high school or college degree, and the average CASI-IRT score was 0.33 (SD: 0.71). There were no meaningful demographic differences across UFP exposure levels, although participants in the highest PNC exposure tertile (as determined by the roadside reference PNC model) had slightly lower CASI-IRT scores compared to those in the lower tertiles. Additional cohort characteristics are reported in Blanco et al. [12].

Model performances and exposure assessment

The reference roadside PNC model had a cross-validated R^2_{MSE} of 0.77. Campaigns that collected measurements during all hours of the day and applied a plume adjustment performed the best, while business hours campaigns generally performed the worst (Fig. 2). Temporal adjustments slightly improved the performance of business hours designs, although they introduced more variability in the results across campaigns, and the resulting improvement was smaller than plume adjustment. Temporally- and plume-adjusted business hours campaigns resulted in exposure models that performed similarly to unadjusted all-

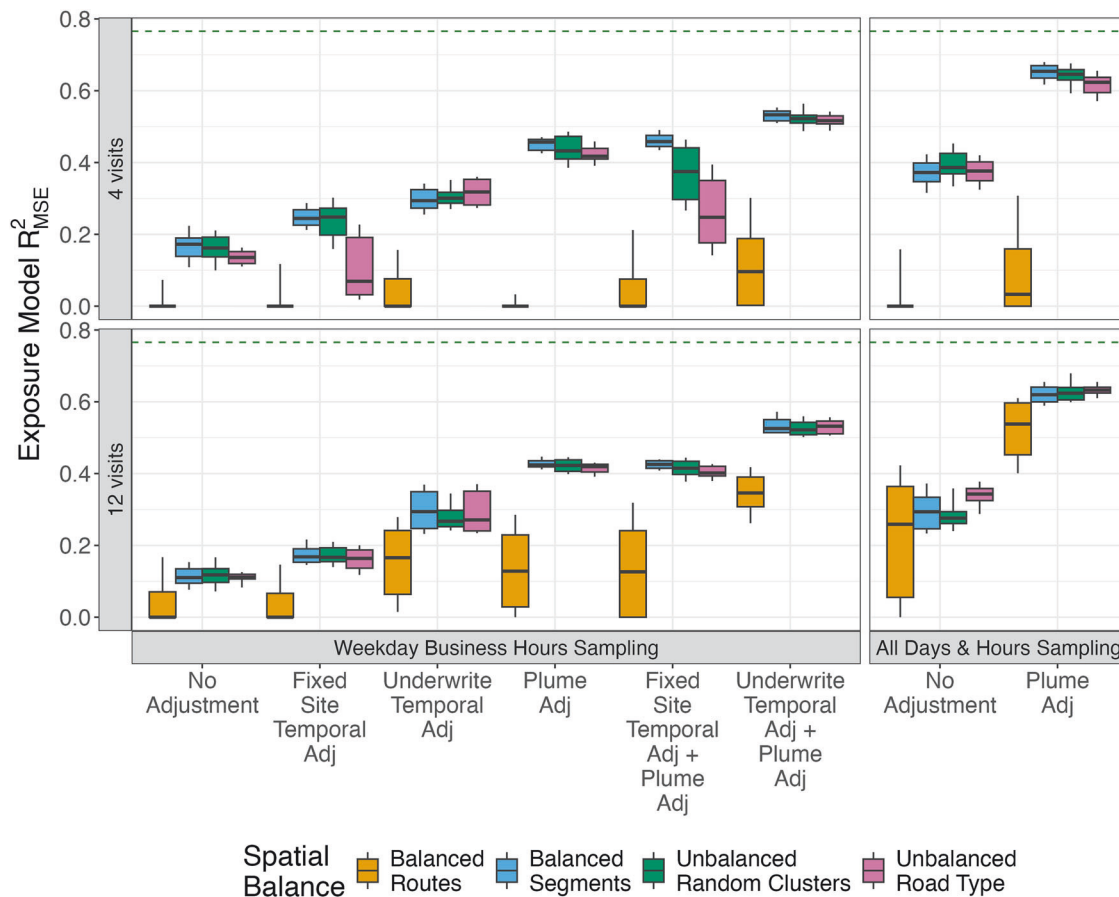


Fig. 2 Out-of-sample PNC (pt/cm³) exposure model performances for on-road campaigns ($N = 30$ campaigns per combination, represented by boxplots). The top plot shows 4-visit campaigns, while the bottom plot shows 12-visit campaigns. R^2_{MSE} is based on the comparison of the predicted PNC at 309 roadside locations to the annual average site estimates at those locations from stationary measures. The dashed green line represents the R^2_{MSE} from the reference roadside exposure model, which is 0.77. Boxes show the median and IQR, and whiskers indicate the 10th and 90th percentiles. Spatially balanced approaches collected an equal number of visits across all locations, whereas unbalanced approaches resulted in uneven sampling, with some locations visited more often than others (e.g., those near certain road types or within spatial clusters). Details on spatial balance are provided in Section 2.4, Restricted Air Pollution Mobile Monitoring Designs.

hours designs. Sampling designs that accounted for route structure—intended to reflect the temporal correlation observed in realized campaigns—consistently performed worse compared to random segment-based sampling, which involves random spatial jumps that are physically unrealistic for actual mobile sampling. In fact, compared to analyses that sampled 4 visits per location at the segment level, analyses that sampled at the route level produced much more variable, worse performances, even when locations are visited 20 times (5 times as much; Fig. S4). Moreover, designs that considered route also benefited from all-hours sampling and the application of temporal and plume adjustments. We found similar results when we used larger spatial clusters to conduct spatially unbalanced sampling (Fig. S5).

As expected, exposure model performances reported using R^2_{reg} were consistently higher than R^2_{MSE} (Fig. S6). Moreover, R^2_{reg} struggled to differentiate performance differences between designs. One exception was the route-based sampling design, which consistently showed poorer performance compared to other designs.

In secondary analyses, we found that fixed-site temporal adjustment sometimes, but not always, improved R^2_{MSE} (Fig. S7). Underwrite temporal adjustments slightly improved R^2_{MSE} .

Cohort predictions from the roadside reference UFP model had a median (IQR) of 7008 (6212–7995). On-road sampling campaigns generated PNC predictions that were slightly more elevated, with plume adjustments reducing these inflated predictions (Fig. S8).

Inferential analyses

In the reduced health model, the adjusted mean baseline CASI-IRT score decreased by 0.021 (95% CI: −0.039, −0.003) per 1900 pt/cm³ increment in PNC, as determined by the reference exposure model. Although associations were slightly attenuated across all on-road designs, the all-day and hourly sampling designs produced the most consistent results, particularly in plume-adjusted campaigns, which produced a median point estimate of −0.015 (IQR: −0.016, −0.014) for the balanced segment designs (Fig. 3). Weekday business hours sampling campaigns resulted in more attenuated health associations (e.g., unadjusted balanced segment design median point estimate −0.008, IQR: −0.009, −0.007). Route-based designs demonstrated greater variability in health estimates (e.g., unadjusted all-day and hour-balanced route design median point estimate −0.015, IQR: −0.018, −0.011). Compared to unadjusted weekday business hours designs, underwrite and plume-adjusted campaigns produced health estimates that were marginally closer to the reference model (e.g., balanced segment design median point estimate −0.010, IQR: −0.011, −0.009). In contrast, fixed-site temporally adjusted designs yielded slightly more attenuated estimates (e.g., balanced segment design median point estimate −0.006, IQR: −0.007, −0.005). Fig. S9 provides point estimates and corresponding 95% confidence intervals for selected designs. Point estimates fluctuate around the reference estimate, while their CIs generally show consistent uncertainty (i.e., similar standard errors).

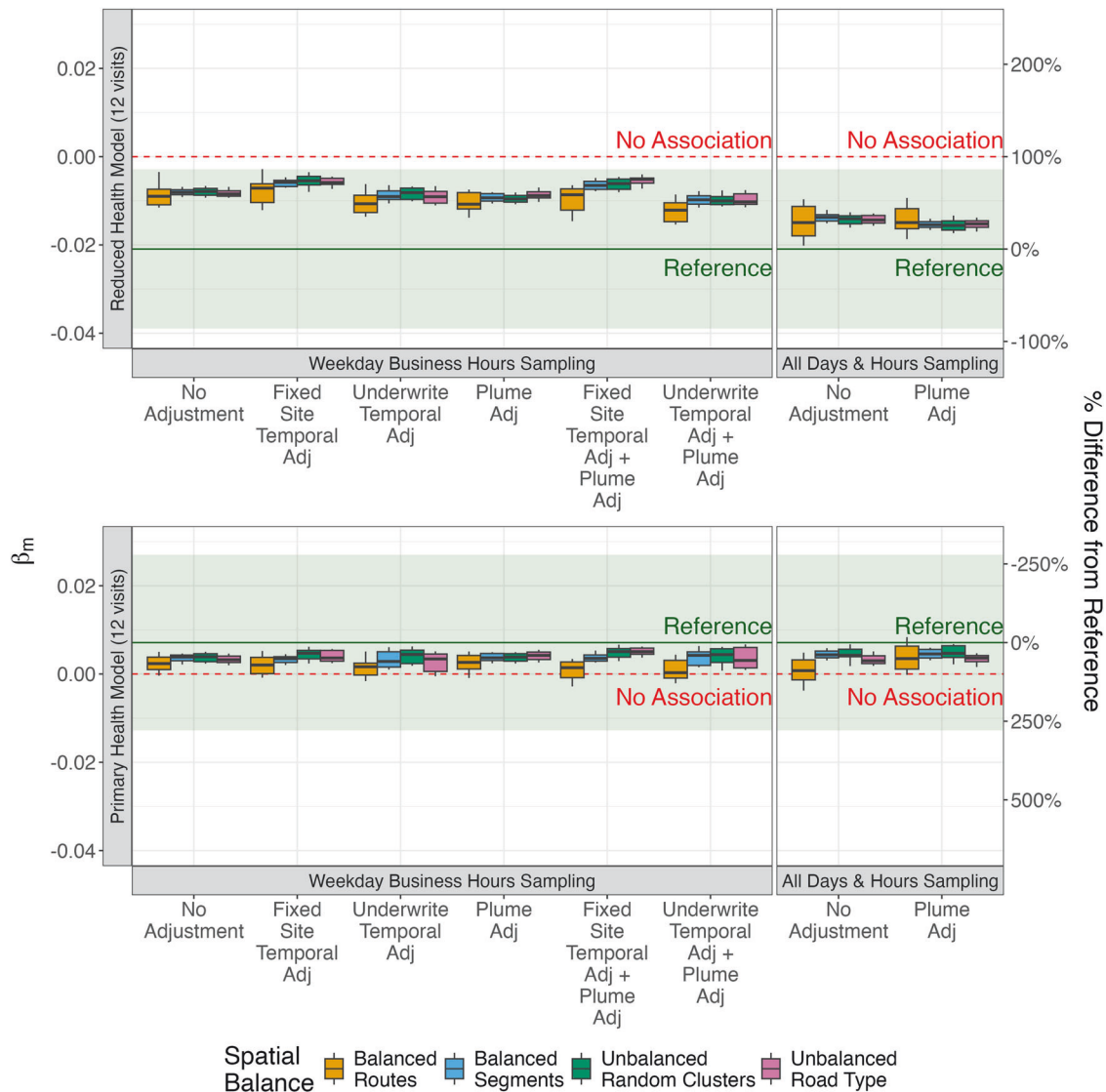


Fig. 3 Estimated association between cognitive function (CASI-IRT) and PNC exposure (per 1900 pt/cm³) for 12-visit campaigns. The top plot shows the reduced health model, while the bottom plot shows the primary health model. The dashed green line and colored area represent the reference health estimate and 95% CI from the roadside exposure model. The dashed red line indicates no association. Boxes show the median and IQR; whiskers illustrate the 10th and 90th percentiles of on-road campaign point estimates. Spatially balanced approaches collected an equal number of visits across all locations, whereas unbalanced approaches resulted in uneven sampling, with some locations visited more often than others (e.g., those near certain road types or within spatial clusters). Details on spatial balance are provided in Section 2.4, Restricted Air Pollution Mobile Monitoring Designs.

In the primary health model, the adjusted mean baseline CASI-IRT score increased by 0.007 (95% CI: −0.013, 0.027) per 1900 pt/cm³ increment in PNC, reflecting associations that were consistent with both adverse and protective effects and not statistically significant. As before, associations were slightly attenuated across all on-road designs, with route-based designs continuing to show greater variability in estimates (e.g., unadjusted all-day and hour-balanced route design median point estimate 0.001, IQR: −0.001, 0.003; vs. similar balanced segment design median point estimate 0.004, IQR: 0.004, 0.005). However, there was substantial overlap in health associations across all designs, making the benefit of specific monitoring designs less clear. Figure S10 provides point estimates and corresponding 95% confidence intervals for selected designs.

We observed similar patterns when we evaluated campaigns with 4 and 12 visits, with 4-visit designs producing more variable results (Fig. S11).

DISCUSSION

We leveraged an extensive mobile monitoring campaign and a long-standing cohort study to gain insight into the impact of common on-road monitoring and analytic decisions on both exposure estimates and health inferences.

Using this reference roadside exposure model in a primary health model, we observed no strong evidence of an association between CASI-IRT score and PNC exposure. Interestingly, this association was slightly attenuated across all on-road monitoring designs, potentially due to lower exposure model performances compared to the reference roadside campaigns (Fig. 2). There was also considerable overlap in the health estimates from all on-road designs, with route-based sampling designs and, as expected, shorter 4 (vs. 12) visit campaigns producing more variable results (Fig. S11), potentially due to the large decreases in temporal representation—which may be important for capturing long-term averages—resulting from these designs. Campaigns that collected spatially and temporally balanced samples had a relatively small

impact in this context; however, this was possibly because health estimates were not statistically significant.

It's notable that, while we aimed to evaluate monitoring design in an analysis with a statistically significant association, we observed this only in reduced health models potentially influenced by residual confounding. In these analyses, all-hours sampling designs with more robust exposure assessment models produced results most consistent with reference health estimates (Fig. 2 and Fig. 3). More restricted weekday business hours sampling, which generally produced less accurate annual average estimates and predicted exposures, produced more attenuated results. Monitoring designs may thus be more critical in contexts when statistically significant associations are present.

Interestingly, while exposure models improved with temporal adjustment, health associations were minimally impacted by temporal adjustments. We previously observed similar results using stationary roadside mobile monitoring data [12]. In this work, we found that collecting temporally balanced visits may be important in some contexts, and that temporally adjusted measurements yield inconsistent improvements in health inferences [10, 12]. Similar findings have been reported in other studies using fixed-site monitoring data, where common temporal adjustment approaches were found to reduce the accuracy of air pollution exposure predictions and even worsen predictions [13]. Moreover, the differing exposure model performances between the fixed-site and underwrite adjustment approaches likely reflect the inherent strengths and limitations of each method. For instance, the fixed-site adjustment approach incorporated measurements across all hours during the study period, including unmonitored times (i.e., overnight), but was constrained to the temporal trend of a single location. Conversely, the underwrite approach leveraged data from all locations within a campaign to provide richer spatial information, but it was limited to the monitoring hours, reducing its ability to correct for non-sampled times. Compared to fixed site adjustments, underwrite adjustments were generally smaller (e.g., IQR: -1087 to 1422 vs. -1406 to 2145 pt/cm³; Table S1). The better performance of the underwrite approach (Fig. S7) may be explained by its ability to capture the spatial variability of UFPs more comprehensively, while requiring less extreme adjustments. It is notable that the fixed-site adjustment approach used in this study was based on a simulated UFP monitoring site rather than actual UFP observations [12]. UFP field monitoring campaigns will face similar constraints given the limited availability of continuous UFP measurements. In cases where exposure predictions improved, temporal adjustments might have reduced Berkson-like measurement error by better aligning the adjusted annual averages with the true reference values. Conversely, in cases where exposure predictions did not improve, the temporal adjustment might have introduced more classical-like measurement error than the reduction in Berkson-like measurement error, ultimately decreasing overall prediction accuracy [50, 51]. Given that these post-hoc adjustments can lead to inconsistent results, conducting temporally balanced sampling increases the likelihood that exposure estimates will be more accurate.

Similar to temporal adjustments, plume adjustment improved the ability of on-road exposure models to predict roadside concentrations—likely due to down-weighting measurements affected by transient spike on-road concentrations—but it did not influence health inferences. This was true both in the primary analyses, where we observed no statistically significant association between UFPs and cognitive function, and in reduced health models, where the association was statistically significant. Plume adjustment may thus be less critical than other monitoring design features—most notably, the inclusion of measurements beyond weekday business hours, which improved health estimates in reduced models. The collection of roadside measurements may also be valuable, given the minimal impact of the plume adjustment applied in this study. We evaluated one approach that leverages both stationary and multi-pollutant measures [14],

but other approaches have been developed to address plumes [15, 25] and could be explored in future work.

Our findings also suggest that route-based sampling should be carefully considered when evaluating reduced sampling campaigns to avoid underestimating the number of visits needed for robust exposure assessment. In practice, adjacent road segments are generally sampled during similar time periods, significantly diminishing the overall temporal coverage that a campaign can achieve. Ignoring this temporal correlation can lead to overly optimistic results, falsely suggesting that far fewer visits are sufficient to produce reliable exposure estimates and health inferences. This may be particularly true when evaluating results with geostatistical models that borrow information from nearby locations.

Interestingly, collecting measurements with equal frequency across locations (i.e., spatial balance) appeared to have the least impact, with minimal effects on both exposure and health models. This is practically helpful for mobile monitoring campaigns, where achieving spatial balance can be logistically difficult—particularly for hard-to-reach locations. In prior work, however, we evaluated more extreme scenarios of spatial imbalance at roadside locations and found that exposure model performance worsened when highly variable sites were sampled only a few times (~2 visits), compared to more stable sites with more frequent sampling (~22 visits) [12]. While this had minimal impact on health inferences on average, it was associated with greater variability across campaigns. Taken together, some degree of spatial imbalance in mobile monitoring may be acceptable; however, it is possible that more extreme imbalances, if evaluated, could result in less stable health estimates.

We found that R^2_{reg} consistently reported much higher exposure model performances and struggled to differentiate between most designs when compared to R^2_{MSE} . While R^2_{reg} was able to detect large performance differences due to restricted sampling hours and route-based sampling approaches, alternative performance metrics like R^2_{MSE} may provide more accurate assessments of the key interest: whether observations and predictions are identical, and not just linearly related. Additionally, it's important to note that we used roadside reference site concentrations, which are more stable than on-road measurements (as discussed in the Methods and below), to evaluate exposure predictions. The choice of reference measures will also impact the results [52].

There are many other monitoring designs and analytic decisions that may be important to consider in future work. Future research may benefit from exploring other sampling strategies, such as approaches that allow for reduced sampling at locations where exposure levels can be more easily predicted or those that evaluate spatially unbalanced route-based sampling. These considerations could further enhance the efficiency and accuracy of mobile monitoring campaigns. Still, findings from our prior work with roadside monitoring campaigns may also be applicable to on-road campaigns [10, 12]. For instance, we previously reported that campaigns with samples collected across 3-4 seasons produced the most robust annual average exposure models and subsequent health inferences. Moreover, although this study focused on UFPs, a pollutant commonly measured with mobile monitoring, we previously observed similar findings for other pollutants, including BC, NO₂, fine particulate matter (PM_{2.5}), and CO₂ when evaluating the number of visits and campaign durations required to produce accurate exposure models [10]. It would be valuable for future work to evaluate these and other design considerations for UFPs, as well as other pollutants across different geographic regions and health outcomes. It is possible that different design approaches may become more important in certain settings, depending on pollutant characteristics and the strength of the health association. Another consideration is that results may vary somewhat when using alternative approaches to exposure modeling (e.g., machine learning). We expect that our

conclusions based on UK-PLS will generalize to other well-performing models, whereas lower-performing approaches (e.g., approaches with limited covariates, or models that have been over-fit) may introduce greater exposure error, potentially obscuring differences in health effect estimates attributable to monitoring design. Finally, while this study focuses on the implications of monitoring design for exposure assessment and health inference, our prior work focuses on a related and practical evaluation of the trade-offs between design, exposure performance, and costs [53].

This study had several limitations. We observed that results varied based not just on the monitoring design and analytic decision, but on the exact health analysis (i.e., primary vs. reduced health models), underscoring the importance of reproducing this work in other settings. Additionally, we used mobile monitoring measurements from 2019 to 2020 in our health analyses under the assumption that exposure surfaces had not changed over time [38–44]. This approach likely introduced exposure measurement error due to temporal misalignment, particularly for earlier years. While our prior work has shown similar findings when restricting analyses to more recent periods [12], future studies should replicate these analyses as longer-term UFP exposure models become available. Furthermore, the plume adjustment applied in this study was derived from the original mobile monitoring campaign, which included a greater number of on-road and roadside measurements than the campaigns sampled in this study. This approach may be overly optimistic, as it assumes that this extensive data will be available. Still, to our knowledge, this is the first study to evaluate the impact of plume adjustment on health inference. Finally, it is possible that design considerations may differ across regions and pollutants (see previous paragraph), as source types, ambient conditions, and dispersion patterns can vary, making some aspects more relevant.

A strength of this study is the use of real data from the long-standing ACT cohort. These data naturally incorporate complexities that might not be captured in a simulation study, thereby enhancing the relevance of our findings. Additionally, we utilized spatially- and temporally balanced roadside measurements to evaluate on-road model performances, which are not always available in practice. Comparing long-term model predictions against more restricted, unbalanced measurements from sampling campaigns can yield unclear results for some designs, particularly when predictions are compared against unstable estimates [10, 52].

In closing, on-road campaigns that collect data using temporally balanced approaches can be used to improve exposure assessment models. Plume and temporal adjustments can sometimes improve exposure models from temporally restricted campaigns, but these may minimally improve subsequent health inferences. Increasing the number of visits per location increases temporal representation and generally leads to more consistent findings. It is important to consider the temporal correlation inherent in realized on-road field campaigns when assessing the impacts of reduced sampling approaches. In terms of health inference implications, we found that on-road monitoring campaigns generally produced attenuated health estimates. In contexts where no association was observed between UFPs and cognition, monitoring design had minimal impact. Monitoring design may have a greater influence in contexts where robust, statistically significant health associations are observed, although more research is needed in this area.

DATA AVAILABILITY

The analytic code for these analyses is available on GitHub (https://github.com/magali17/hei_aim3a). A Zenodo repository currently hosts relevant data from the mobile monitoring campaign (<https://zenodo.org/records/13761282>). The authors plan to add the on-road monitoring data to this repository following the completion of related ongoing analyses.

REFERENCES

- Kim SY, Blanco MN, Bi J, Larson TV, Sheppard L. Exposure assessment for air pollution epidemiology: a scoping review of emerging monitoring platforms and designs. *Environ Res.* 2023;223:115451.
- Brugge D, Fuller CH, editors. *Ambient combustion ultrafine particles and health*. New York: Nova Science Publishers; 2020. 436 p.
- Flood-Garibay JA, Angulo-Molina A, Méndez-Rojas MÁ. Particulate matter and ultrafine particles in urban air pollution and their effect on the nervous system. *Environ Sci Process Impacts.* 2023;25:704–26.
- HEI. Understanding the Health Effects of Ambient Ultrafine Particles. Perspectives 3. 2013. Available from: <https://www.healtheffects.org/publication/understanding-health-effects-ambient-ultrafine-particles>.
- US EPA. Integrated science assessment (ISA) for particulate matter. US Environmental Protection Agency; Washington: 2019. Available from: <https://cfpub.epa.gov/ncea/isa/recordisplay.cfm?deid=347534>.
- Messier KP, Chambliss SE, Gani S, Alvarez R, Brauer M, Choi JJ, et al. Mapping air pollution with Google Street View cars: efficient approaches with mobile monitoring and land use regression. *Environ Sci Technol.* 2018;52:12563–72.
- Kerckhoffs J, Hoek G, Vermeulen R. Mobile monitoring of air pollutants; performance evaluation of a mixed-model land use regression framework in relation to the number of drive days. *Environ Res.* 2024;240:117457.
- Hatzopoulou M, Valois MF, Levy I, Mihele C, Lu G, Bagg S, et al. Robustness of land-use regression models developed from mobile air pollutant measurements. *Environ Sci Technol.* 2017;51:3938–47.
- Upadhyay AR, Kushwaha M, Agrawal P, Gingrich JD, Asundi J, Sreekanth V, et al. Multi-season mobile monitoring campaign of on-road air pollution in Bengaluru, India: high-resolution mapping and estimation of quasi-emission factors. *Sci Total Environ.* 2024;914:169987.
- Blanco MN, Bi J, Austin E, Larson TV, Marshall JD, Sheppard L. Impact of mobile monitoring network design on air pollution exposure assessment models. *Environ Sci Technol.* 2023;57:440–50.
- Blanco MN, Doubleday A, Austin E, Marshall JD, Seto E, Larson TV, et al. Design and evaluation of short-term monitoring campaigns for long-term air pollution exposure assessment. *J Expo Sci Environ Epidemiol.* 2023;33:465–73.
- Blanco MN, Szpiro AA, Crane PK, Sheppard L. Ultrafine particles and late-life cognitive function: Influence of stationary mobile monitoring design on health inferences. *Environ Pollut.* 2025;374:126222.
- Chastko K, Adams M. Assessing the accuracy of long-term air pollution estimates produced with temporally adjusted short-term observations from unstructured sampling. *J Environ Manag.* 2019;240:249–58.
- Doubleday A, Blanco MN, Austin E, Marshall JD, Larson TV, Sheppard L. Characterizing ultrafine particle mobile monitoring data for epidemiology. *Environ Sci Technol.* 2023;57:9538–47.
- Kerckhoffs J, Hoek G, Messier KP, Brunekreef B, Meliefste K, Klompmaier JO, et al. Comparison of ultrafine particle and black carbon concentration predictions from a mobile and short-term stationary land-use regression model. *Environ Sci Technol.* 2016;50:12894–902.
- Kerckhoffs J, Hoek G, Vlaanderen J, van Nunen E, Messier K, Brunekreef B, et al. Robustness of intra-urban land-use regression models for ultrafine particles and black carbon based on mobile monitoring. *Environ Res.* 2017;159:500–8.
- Minet L, Liu R, Valois MF, Xu J, Weichenthal S, Hatzopoulou M. Development and comparison of air pollution exposure surfaces derived from on-road mobile monitoring and short-term stationary sidewalk measurements. *Environ Sci Technol.* 2018;52:3512–9.
- Chambliss SE, Preble CV, Caubel JJ, Cados T, Messier KP, Alvarez RA, et al. Comparison of mobile and fixed-site black carbon measurements for high-resolution urban pollution mapping. *Environ Sci Technol.* 2020;54:7848–57.
- Johnson J, Kittelson D, Watts W. Source apportionment of diesel and spark ignition exhaust aerosol using on-road data from the Minneapolis metropolitan area. *Atmos Environ.* 2005;39:2111–21.
- Kelp M, Gould T, Austin E, Marshall JD, Yost M, Simpson C, et al. Sensitivity analysis of area-wide, mobile source emission factors to high-emitter vehicles in Los Angeles. *Atmos Environ.* 2020;223:117212.
- May AA, Nguyen NT, Presto AA, Gordon TD, Lipsky EM, Karve M, et al. Gas- and particle-phase primary emissions from in-use, on-road gasoline and diesel vehicles. *Atmos Environ.* 2014;88:247–60.
- Tong Z, Li Y, Lin Q, Wang H, Zhang S, Wu Y, et al. Uncertainty investigation of the plume-chasing method for measuring on-road NOx emission factors of heavy-duty diesel vehicles. *J Hazard Mater.* 2022;424:127372.
- Wang X, Westerdahl D, Chen LC, Wu Y, Hao J, Pan X, et al. Evaluating the air quality impacts of the 2008 Beijing Olympic Games: on-road emission factors and black carbon profiles. *Atmos Environ.* 2009;43:4535–43.
- Wen Y, Wang H, Larson T, Kelp M, Zhang S, Wu Y, et al. On-highway vehicle emission factors and spatial patterns, based on mobile monitoring and absolute principal component score. *Sci Total Environ.* 2019;676:242–51.

25. Yuan Z, Kerckhoffs J, Hoek G, Vermeulen R. A knowledge transfer approach to map long-term concentrations of hyperlocal air pollution from short-term mobile measurements. *Environ Sci Technol*. 2022;56:13820–8.
26. Blanco MN, Gasset A, Gould T, Doubleday A, Slager DL, Austin E, et al. Characterization of annual average traffic-related air pollution concentrations in the greater Seattle area from a year-long mobile monitoring campaign. *Environ Sci Technol*. 2022;56:11460–72.
27. Kukull WA, Higdon R, Bowen JD, McCormick WC, Teri L, Schellenberg GD, et al. Dementia and Alzheimer's disease incidence: a prospective cohort study. *Arch Neurol*. 2002;59:1737–46.
28. Teng EL, Hasegawa K, Homma A, Imai Y, Larson E, Graves A, et al. The Cognitive Abilities Screening Instrument (CASI): a practical test for cross-cultural epidemiological studies of dementia. *Int Psychogeriatr*. 2004;6:45–58.
29. Crane PK, Narasimhalu K, Gibbons LE, Mungas DM, Haneuse S, Larson EB, et al. Item response theory facilitated calibrating cognitive tests and reduced bias in estimated rates of decline. *J Clin Epidemiol*. 2008;61:1018–27.
30. Ehlenbach WJ, Hough CL, Crane PK, Haneuse S, Carson SS, Curtis JR, et al. Association between acute care and critical illness hospitalization and cognitive function in older adults. *JAMA*. 2010;303:763–70.
31. Li G, Larson EB, Shofer JB, Crane PK, Gibbons LE, McCormick W, et al. Cognitive trajectory changes over 20 years before dementia diagnosis: a large cohort study. *J Am Geriatr Soc*. 2017;65:2627–33.
32. Eeftens M, Tsai MY, Ampe C, Anwander B, Beelen R, Bellander T, et al. Spatial variation of PM_{2.5}, PM₁₀, PM_{2.5} absorbance and PM_{coarse} concentrations between and within 20 European study areas and the relationship with NO₂—results of the ESCAPE project. *Atmos Environ*. 2012;62:303–17.
33. Hoek G, Meliefste K, Cyrys J, Lewné M, Bellander T, Brauer M, et al. Spatial variability of fine particle concentrations in three European areas. *Atmos Environ*. 2002;36:4077–88.
34. Montagne DR, Hoek G, Klompmaker JO, Wang M, Meliefste K, Brunekreef B. Land use regression models for ultrafine particles and black carbon based on short-term monitoring predict past spatial variation. *Environ Sci Technol*. 2015;49:8712–20.
35. van Nunen E, Vermeulen R, Tsai MY, Probst-Hensch N, Ineichen A, Davey M, et al. Land use regression models for ultrafine particles in six European areas. *Environ Sci Technol*. 2017;51:3336–45.
36. Hankey S, Marshall JD. On-bicycle exposure to particulate air pollution: particle number, black carbon, PM_{2.5}, and particle size. *Atmos Environ*. 2015;122:65–73.
37. Hankey S, Marshall JD. Land use regression models of on-road particulate air pollution (particle number, black carbon, PM_{2.5}, particle Size) using mobile monitoring. *Environ Sci Technol*. 2015;49:9194–202.
38. Blanco MN. Traffic-related air pollution and dementia incidence in a Seattle-based, prospective cohort study. Dissertation, University of Washington, 2021.
39. Blanco MN, Shaffer RM, Li G, Adar SD, Carone M, Szpiro AA, et al. Traffic-related air pollution and dementia incidence in the Adult Changes in Thought Study. *Environ Int*. 2024;183.
40. Kim SY, Olives C, Sheppard L, Sampson PD, Larson TV, Keller JP, et al. Historical prediction modeling approach for estimating long-term concentrations of PM_{2.5} in cohort studies before the 1999 implementation of widespread monitoring. *Environ Health Perspect*. 2017;125:38–46.
41. Levy I, Levin N, Yuval Y, Schwartz JD, Kark JD. Back-extrapolating a land use regression model for estimating past exposures to traffic-related air pollution. *Environ Sci Technol*. 2015;49:3603–10.
42. Meng J, Li C, Martin RV, van Donkelaar A, Hystad P, Brauer M. Estimated long-term (1981–2016) concentrations of ambient fine particulate matter across North America from chemical transport modeling, satellite remote sensing, and ground-based measurements. *Environ Sci Technol*. 2019;53:5071–9.
43. Molter A, Lindley S, de Vocht F, Simpson A, Agius R. Modelling air pollution for epidemiologic research—Part II: predicting temporal variation through land use regression. *Sci Total Environ*. 2010;409:211–7.
44. Wang Y, Hopke PK, Chalupa DC, Utell MJ. Long-term study of urban ultrafine particles and other pollutants. *Atmos Environ*. 2011;45:7672–80.
45. Miles JN, Weden MM, Lavery D, Escarce JJ, Cagney KA, Shih RA. Constructing a time-invariant measure of the socio-economic status of U.S. census tracts. *J Urban Health*. 2016;93:213–32.
46. Zeileis A. Econometric computing with HC and HAC covariance matrix estimators. *J Stat Soft*. 2004;11. Available from: <http://www.jstatsoft.org/v11/i10/>.
47. Zeileis A, Köll S, Graham N. Various versatile variances: an object-oriented implementation of clustered covariances in R. *J Stat Soft*. 2020;95. Available from: <http://www.jstatsoft.org/v95/i01/>.
48. Zeileis A, Hothorn T. Diagnostic checking in regression relationships. *R N*. 2002;2:7–10.
49. R Core Team R: A language and environment for statistical computing. 2023. Available from: <https://www.R-project.org>.
50. Szpiro AA, Sheppard L, Lumley T. Efficient measurement error correction with spatially misaligned data. *Biostatistics*. 2011;12:610–23.
51. Szpiro AA, Paciorek CJ. Measurement error in two-stage analyses, with application to air pollution epidemiology. *Environmetrics*. 2013;24:501–17.
52. Blanco MN, Doubleday A, Austin E, Marshall JD, Seto E, Larson TV, et al. Design and evaluation of short-term monitoring campaigns for long-term air pollution exposure assessment. *J Expo Sci Environ Epidemiol*. 2022;33:465–73.
53. Kim SY, Gasset AJ, Blanco MN, Sheppard L. Ultrafine particle mobile monitoring study designs for epidemiology: Cost and performance comparisons. *Environ Health Perspect*. 2025;133:47010.

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.

AUTHOR CONTRIBUTIONS

MB: Formal Analysis, Methodology, Data Curation, Investigation, Writing—original draft. AD: Data Curation, Writing—review and editing. AAS: Investigation, Methodology, Writing—review and editing. JDM: Investigation, Methodology, Writing—review and editing. PC: Investigation, Resources, Writing—review and editing. LS: Conceptualization, Funding Acquisition, Methodology, Supervision, Writing—review and editing

COMPETING INTERESTS

The authors declare no competing interests.

ETHICAL APPROVAL AND CONSENT TO PARTICIPATE

All methods were performed in accordance with the relevant guidelines and regulations. Study procedures were approved by the University of Washington (52189, STUDY00009108) and Kaiser Permanente (278254 - Repository for ADPR/ACT Study) institutional review boards. Informed consent was obtained from all participants.

ADDITIONAL INFORMATION

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41370-026-00845-y>.

Correspondence and requests for materials should be addressed to Magali N. Blanco.

Reprints and permission information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons

Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2026

Supplemental Material

Influence of on-road mobile monitoring design on ultrafine particle exposure models and cognitive health inferences

Magali N. Blanco,¹ Annie Doubleday,¹ Adam A. Szpiro,² Julian D. Marshall,³ Paul K. Crane,⁴ Lianne Sheppard^{1,2}

¹Department of Environmental and Occupational Health Sciences, University of Washington, Seattle, WA USA

²Department of Biostatistics, University of Washington, Seattle, WA USA

³Department of Civil & Environmental Engineering, University of Washington, Seattle, WA USA

⁴Department of Medicine, University of Washington, Seattle, WA, USA

Table of Contents

1	METHODS.....	4
2	RESULTS.....	8

List of Figures

FIGURE S1. SPATIAL CLUSTERS USED TO CONDUCT THE UNBALANCED SAMPLING DESIGNS. THE MAIN CLUSTERS (LEFT) INCLUDE AN AVERAGE OF 93 100 M SEGMENTS PER CLUSTER. SENSITIVITY CLUSTERS HAVE APPROXIMATELY THREE TIMES AS MANY SEGMENTS.....	4
FIGURE S2. EXAMPLE TIME SERIES OF MEASURED AND SIMULATED BACKGROUND PNC FOR TWO SAMPLING DAYS. ONE-SECOND P-TRAK FIELD MEASUREMENTS ARE SHOWN IN PURPLE. BACKGROUND PNC IS CALCULATED AS THE 1ST PERCENTILE OF THREE-HOUR ROLLING WINDOWS (GREEN); THESE VALUES ARE USED TO ESTIMATE HOURLY PNC (RED) AND LONG-TERM AVERAGE PNC (TEAL). THE PRIMARY ANALYSIS FOCUSES ON NON-HIGHWAY MEASUREMENTS, WHILE SENSITIVITY ANALYSES INCORPORATE ALL ROAD TYPES, REPRESENTED BY DIFFERENT LINE STYLES.....	6
FIGURE S3. EXAMPLE OF TEMPORAL ADJUSTMENTS IN SENSITIVITY ANALYSES EVALUATING DIFFERENT PERCENTILES (1 ST TO 10 TH) AND TIME WINDOWS (1 TO 3 HOURS) TO DEFINE UNDERWRITE FUNCTIONS. (A) THE TOP TWO ROWS SHOW ESTIMATED HOURLY BACKGROUND CONCENTRATIONS AND LONG-TERM AVERAGES (HORIZONTAL LINES) FOR DIFFERENT SAMPLING DAYS. (B) THE BOTTOM TWO ROWS SHOW THE CORRESPONDING TEMPORAL ADJUSTMENTS. DIFFERENT ROAD TYPES ("ALL ROAD TYPES" VS. "NO HIGHWAYS") ARE REPRESENTED BY DIFFERENT LINE TYPES. DESPITE VARIATIONS IN TIME WINDOWS, PERCENTILES, AND ROAD TYPES, RESULTS ARE HIGHLY CONSISTENT ACROSS ALL SCENARIOS.....	7
FIGURE S4. OUT-OF-SAMPLE PNC (PT/CM ³) EXPOSURE MODEL PERFORMANCES FOR ON-ROAD CAMPAIGNS (N=30 CAMPAIGNS PER COMBINATION - I.E., BOXPLOT) WHEN SAMPLING AT THE SEGMENT- VS. ROUTE-LEVEL. COMPARED TO ANALYSES THAT SAMPLE 4 VISITS PER LOCATION AT THE SEGMENT LEVEL, ANALYSES THAT CONSIDER LOGISTICAL FIELD CONSTRAINTS AND SAMPLE ENTIRE ROUTES GENERALLY PERFORM WORSE, EVEN WHEN LOCATIONS ARE VISITED 20 TIMES (5 TIMES AS MUCH). <i>RMSE2</i> IS BASED ON A COMPARISON OF THE PREDICTED PNC AT 309 ROADSIDE LOCATIONS TO THE ANNUAL AVERAGE SITE ESTIMATES AT THOSE LOCATIONS FROM STATIONARY MEASURES. THE BOXES ILLUSTRATE THE MEDIAN AND IQR, WHILE THE WHISKERS EXTEND TO THE 10TH AND 90TH PERCENTILES. THE DASHED LINE SHOWS THE PERFORMANCE OF THE STATIONARY ROADSIDE MODEL.	8
FIGURE S5. OUT-OF-SAMPLE PNC (PT/CM ³) EXPOSURE MODEL PERFORMANCES FOR ON-ROAD CAMPAIGNS (N=30 CAMPAIGNS PER COMBINATION - I.E., BOXPLOT) BASED ON SPATIAL CLUSTERS USED IN THE MAIN ANALYSIS AND ALTERNATIVE, LARGER CLUSTERS. RESULTS ARE FOR CAMPAIGNS WITH 12 VISITS PER LOCATION. <i>RMSE2</i> IS BASED ON A COMPARISON OF THE PREDICTED PNC AT 309 ROADSIDE LOCATIONS TO THE ANNUAL AVERAGE SITE ESTIMATES AT THOSE LOCATIONS FROM STATIONARY MEASURES. THE BOXES ILLUSTRATE THE MEDIAN AND IQR, WHILE THE WHISKERS EXTEND TO THE 10TH AND 90TH PERCENTILES. THE DASHED LINE SHOWS THE PERFORMANCE OF THE STATIONARY ROADSIDE MODEL.	9
FIGURE S6. OUT-OF-SAMPLE PNC (PT/CM ³) EXPOSURE MODEL PERFORMANCES FOR ON-ROAD CAMPAIGNS (N=30 CAMPAIGNS PER COMBINATION - I.E., BOXPLOT) EVALUATED USING <i>Rreg2</i> FOR 12-VISIT CAMPAIGNS. <i>Rreg2</i> IS BASED ON A COMPARISON OF THE PREDICTED PNC AT 309 STATIONARY LOCATIONS AND THE ANNUAL AVERAGE SITE ESTIMATE BASED ON STATIONARY MEASURES. THE BOXES ILLUSTRATE THE MEDIAN AND IQR, WHILE THE WHISKERS EXTEND TO THE 10TH AND 90TH PERCENTILES. THE DASHED LINE SHOWS THE PERFORMANCE OF THE STATIONARY ROADSIDE MODEL FOR RESPECTIVE R ² , WHICH IS THE SAME FOR <i>RMSE2</i> AND <i>Rreg2</i>	10
FIGURE S7. COMPARISON OF <i>RMSE2</i> FOR A SUBSET OF 12-VISIT BUSINESS HOURS SAMPLING DESIGNS WITH AND WITHOUT TEMPORAL ADJUSTMENT. THE DIAGONAL DASHED LINE IS THE 1-1 LINE; VALUES ABOVE THIS LINE INDICATE BETTER RESULTING <i>RMSE2</i> VALUES WHEN COMPARED TO UNADJUSTED CAMPAIGNS.....	11
FIGURE S8. PREDICTED FIVE-YEAR AVERAGE PNC EXPOSURE FOR ACT PARTICIPANTS ACROSS 30 SAMPLING CAMPAIGNS FOR EACH DESIGN (I.E., BOXPLOT). THE GREEN DASHED LINE AND AREA REPRESENT THE MEDIAN AND IQR OF CROSS-VALIDATED PNC PREDICTIONS FROM THE REFERENCE ALL-DATA STATIONARY EXPOSURE MODEL. THE BOXES ILLUSTRATE THE MEDIAN AND IQR, WHILE THE WHISKERS EXTEND TO THE 10TH AND 90TH PERCENTILES.	12
FIGURE S9. POINT ESTIMATES AND 95% CONFIDENCE INTERVALS (CI) FROM REDUCED HEALTH MODELS OF THE ADJUSTED ASSOCIATION BETWEEN PNC (1,900 PT/CM ³) AND CASI-IRT FOR SELECTED CAMPAIGNS. THE GREEN LINE AND SHADED AREA REPRESENT THE POINT ESTIMATE AND 95% CI FROM THE ROADSIDE ALL-DATA EXPOSURE MODEL, WHILE THE RED LINE INDICATES NO ASSOCIATION.	13

FIGURE S10. POINT ESTIMATES AND 95% CONFIDENCE INTERVALS (CI) FROM PRIMARY HEALTH MODELS OF THE ADJUSTED ASSOCIATION BETWEEN PNC (1,900 PT/CM ³) AND CASI-IRT FOR SELECTED CAMPAIGNS. THE GREEN LINE AND SHADED AREA REPRESENT THE POINT ESTIMATE AND 95% CI FROM THE ROADSIDE ALL-DATA EXPOSURE MODEL, WHILE THE RED LINE INDICATES NO ASSOCIATION.	14
FIGURE S11. ESTIMATED ASSOCIATION BETWEEN COGNITIVE FUNCTION (CASI-IRT) AND PNC EXPOSURE (PER 1,900 PT/CM ³) FOR 4-VISIT CAMPAIGNS. THE DASHED GREEN LINE AND COLORED AREA REPRESENT THE REFERENCE HEALTH ESTIMATE AND 95% CI FROM THE ROADSIDE EXPOSURE MODEL. THE DASHED RED LINE INDICATES NO ASSOCIATION. BOXES SHOW THE MEDIAN AND IQR; WHISKERS ILLUSTRATE THE 10TH AND 90TH PERCENTILES OF ON-ROAD CAMPAIGN POINT ESTIMATES.	15

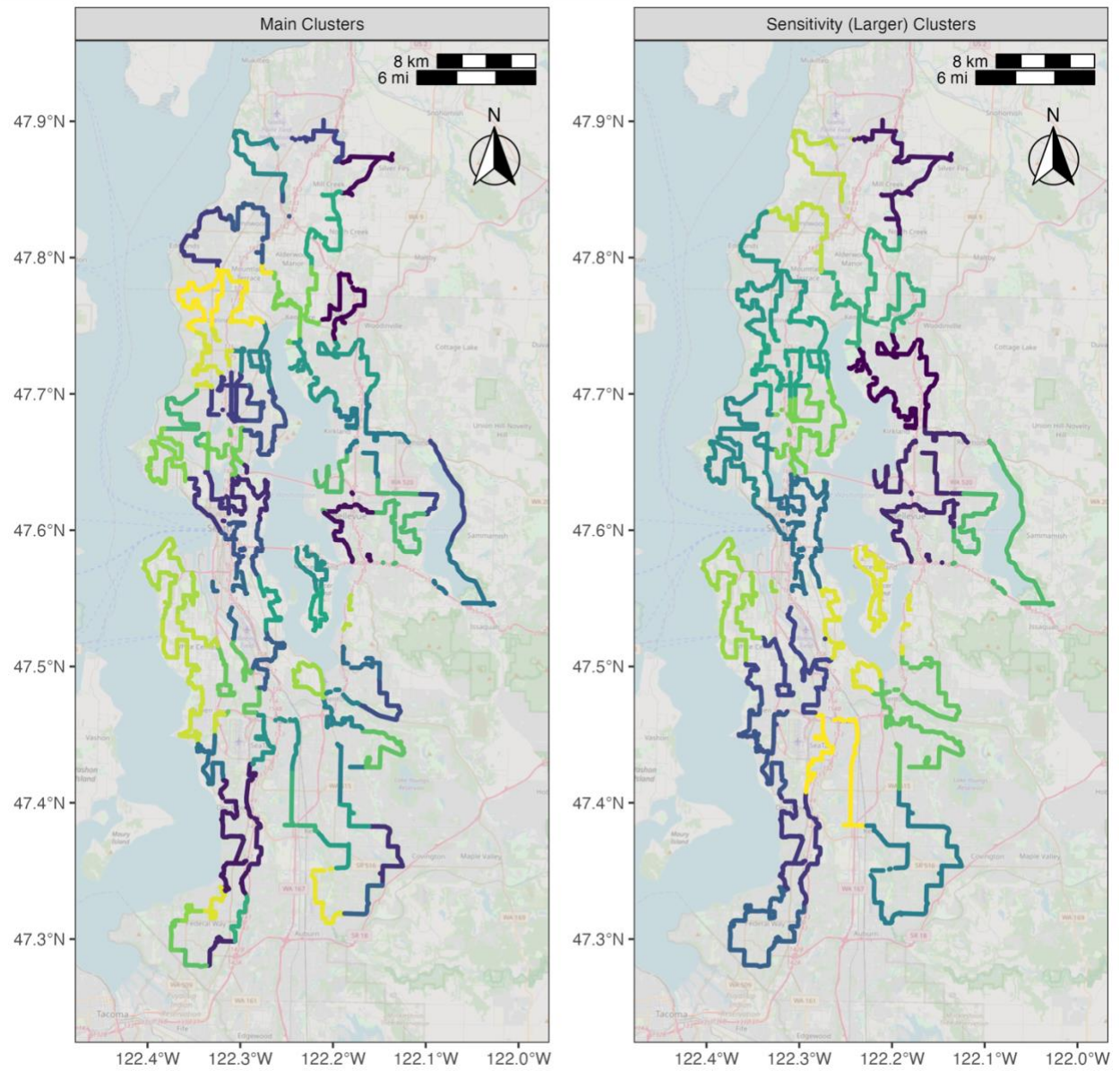
LIST OF TABLES

TABLE S1. DISTRIBUTION OF HOURLY P-TRAK PNC (PT/CM ³) TEMPORAL ADJUSTMENTS APPLIED TO ADJUST BUSINESS HOURS CAMPAIGN VISITS.	8
---	---

List of Notes

NOTE S1. SUMMARY OF TEMPORAL ADJUSTMENT USING AN UNDERWRITE FUNCTION APPROACH	5
---	---

1 Methods



Map data © OpenStreetMap contributors

Figure S1. Spatial clusters used to conduct the unbalanced sampling designs. The main clusters (left) include an average of 93 100 m segments per cluster. Sensitivity clusters have approximately three times as many segments.

In summary, our temporal approach involved the following steps for each sampling campaign:

1. Simulating background PNC levels using the collected 1-second PNC measurements to calculate the first percentile of three-hour rolling averages.
2. Calculating hourly and long-term (~1 year) average background PNC.
3. Estimating hourly temporal adjustments based on the difference between hourly and long-term average background PNC.
4. Applying temporal adjustments to mobile monitoring visits conducted during business hours.
5. Estimating temporally adjusted annual average locations concentrations.

First, we simulated a continuous background PNC monitoring site using the collected PNC data since one was otherwise unavailable. We started with 3,359,505 unique 1-second on-road P-TRAK PNC measurements during the mobile monitoring campaign. To reduce the influence of extreme and potentially less reliable readings, we winsorized these measurement by setting measurements below the 5th percentile and above the 95th percentile to those respective thresholds. Additionally, we excluded measurements taken on highways (A1 roads) since they are less likely to represent background concentrations. This processing resulted in 2,978,809 (89%) 1-second measurements used to estimate background PNC over time, defined as the first percentile concentration of three-hour rolling windows (see Figure S2). Rolling windows were centrally aligned (e.g., for a three-hour window, data from 1.5 hours before and after were used). In sensitivity analyses, we evaluated different quantile-window combinations and included highway readings.

We used the resulting 1-second PNC background estimates, x_s (where $s = \{1, \dots, 2,978,809\}$) during the study period (~1 year) to calculate: (a) background PNC levels for each sampled hour h (where $h = \{1, \dots, 1,571\}$), and (b) a long-term average (LTA) background PNC (μ_{LTA} ; Figure S3):

$$\mu_h = \frac{1}{S} \sum_{s=1}^S x_{s,h} \quad (S1)$$

$$\mu_{LTA} = \frac{1}{S} \sum_{s=1}^S x_s \quad (S2)$$

We then calculated hourly adjustment factors, δ_h , defined as the difference between the long-term average and hourly averages (see Table S1 for distribution of adjustments):

$$\delta_h = \mu_{LTA} - \mu_h \quad (S3)$$

These temporal adjustments were applied to adjust PNC measurements based on the exact time of sampling, allowing us to calculate temporally adjusted annual average PNC for each road segment. The temporally adjusted PNC for each visit v and location l is given by:

$$PNC_{v,l,\delta} = PNC_{v,l} + \delta_h \quad (S4)$$

The temporally adjusted long-term average PNC for each location l is then:

$$PNC_{LTA,l,\delta} = \sum_{v=1}^{V_l} PNC_{v,l,\delta} \quad (S5)$$

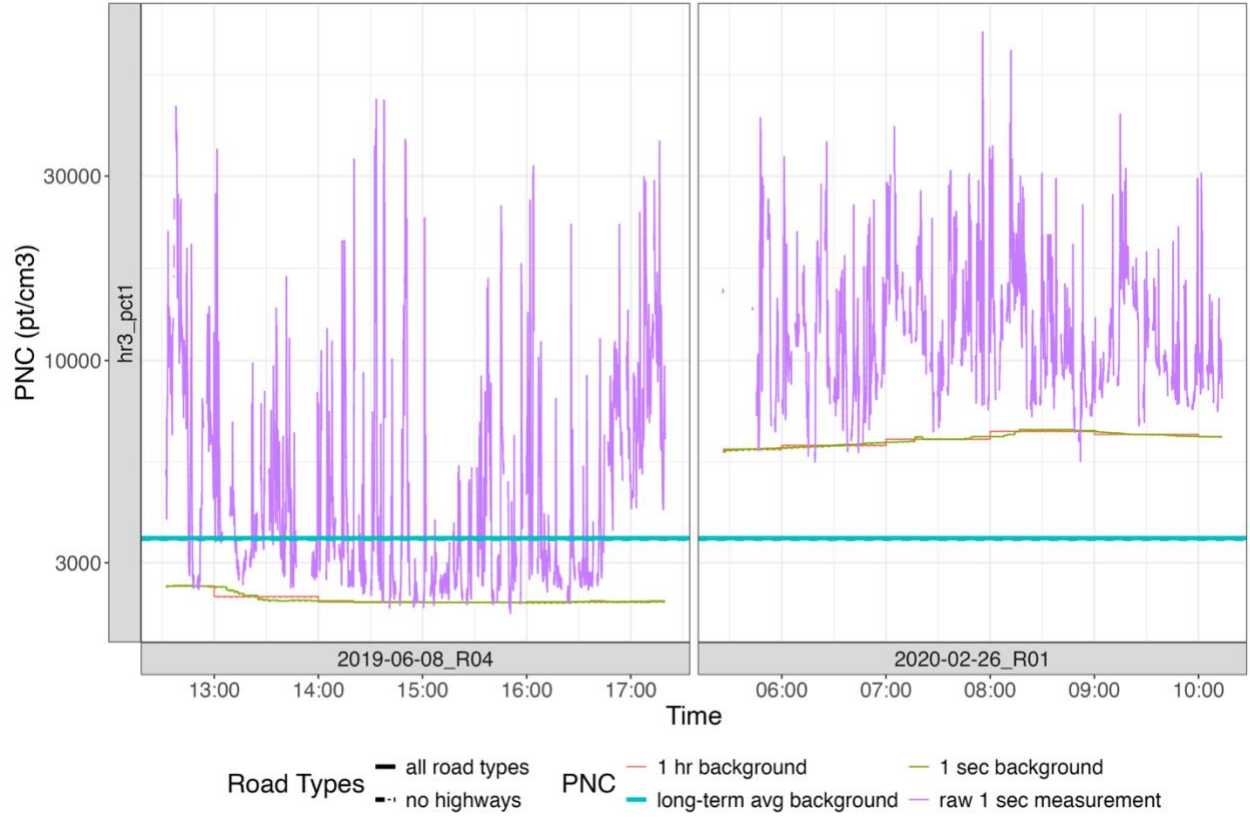
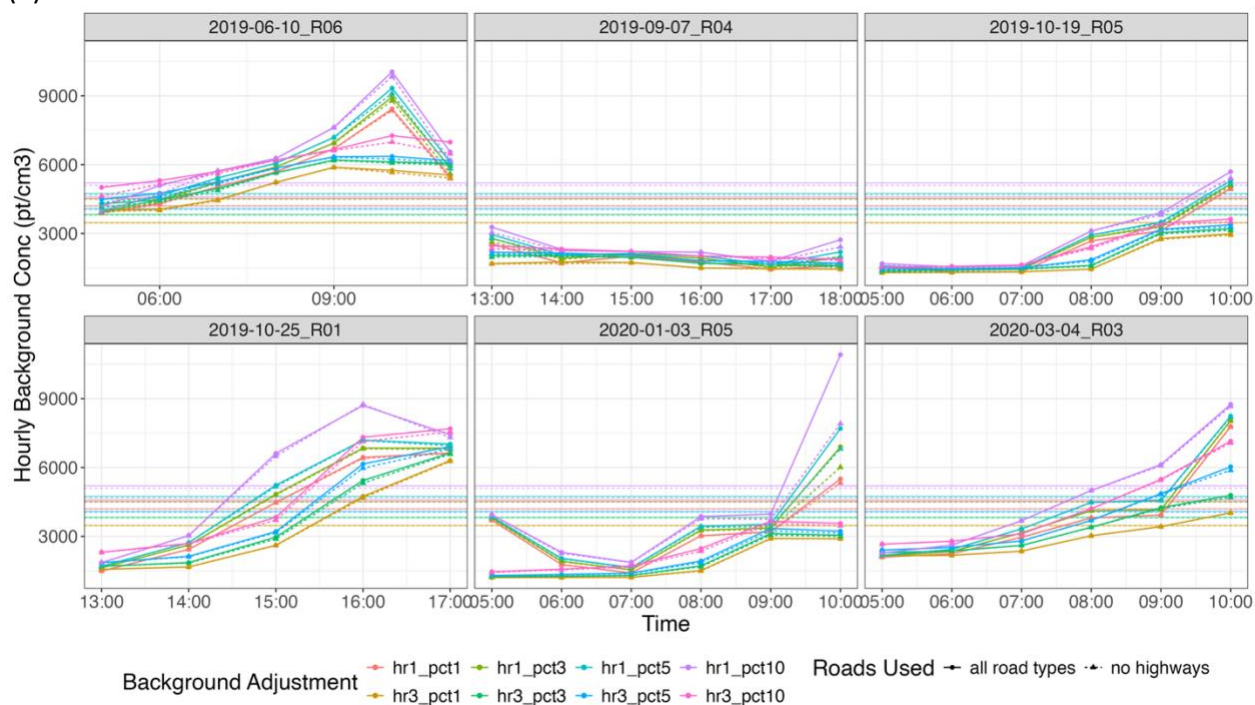


Figure S2. Example time series of measured and simulated background PNC for two sampling days. One-second P-TRAK field measurements are shown in purple. Background PNC is calculated as the 1st percentile of three-hour rolling windows (green); these values are used to estimate hourly PNC (red) and long-term average PNC (teal). The primary analysis focuses on non-highway measurements, while sensitivity analyses incorporate all road types, represented by different line styles.

(a)



(b)

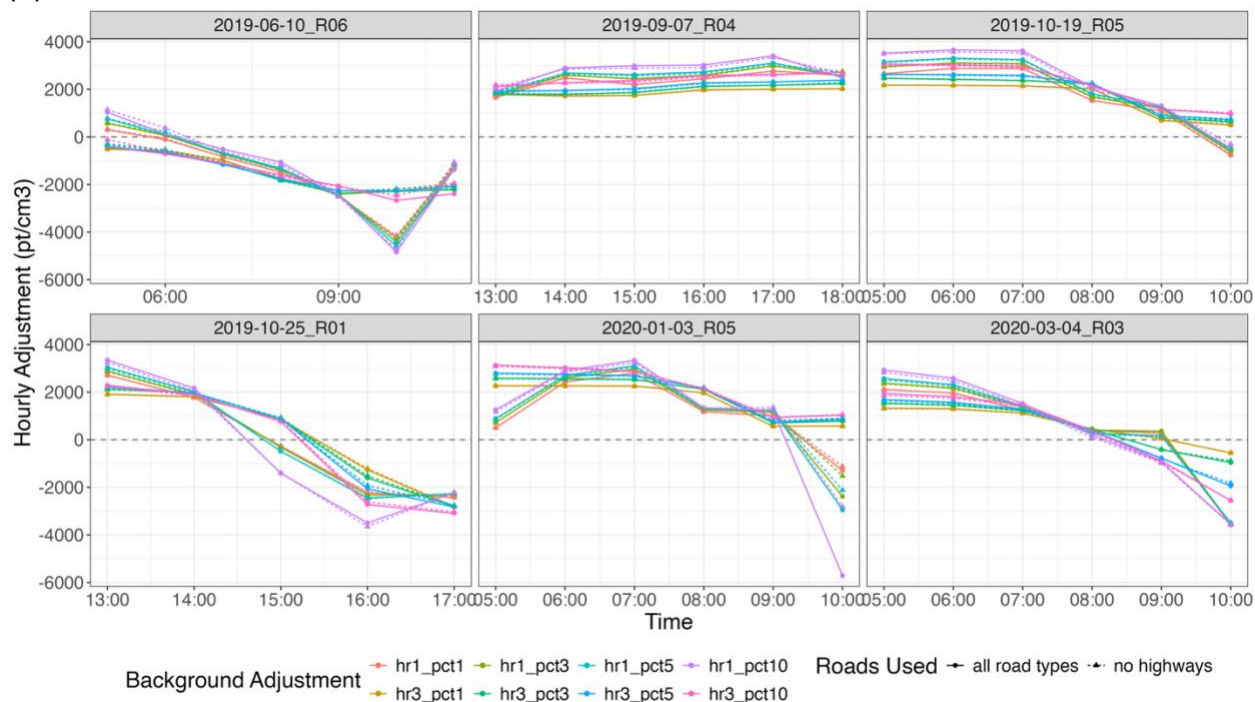


Figure S3. Example of temporal adjustments in sensitivity analyses evaluating different percentiles (1st to 10th) and time windows (1 to 3 hours) to define underwrite functions. (a) The top two rows show estimated hourly background concentrations and long-term averages (horizontal lines) for different sampling days. (b) The bottom two rows show the corresponding temporal adjustments. Different road types ("all road types" vs. "no highways") are represented by different line types. Despite variations in time windows, percentiles, and road types, results are highly consistent across all scenarios.

Table S1. Distribution of hourly P-TRAK PNC (pt/cm³) temporal adjustments applied to adjust business hours campaign visits.

Adjustment	Hours	Min	Q25	Median	Mean	Q75	Max	SD	IQR
Fixed Site	9,358	-5,592	-1,406	641	131	2,145	3,544	2,537	3,550
Underwrite	1,571	-11,387	-1,087	484	-53	1,422	2,972	2,054	2,509

2 Results

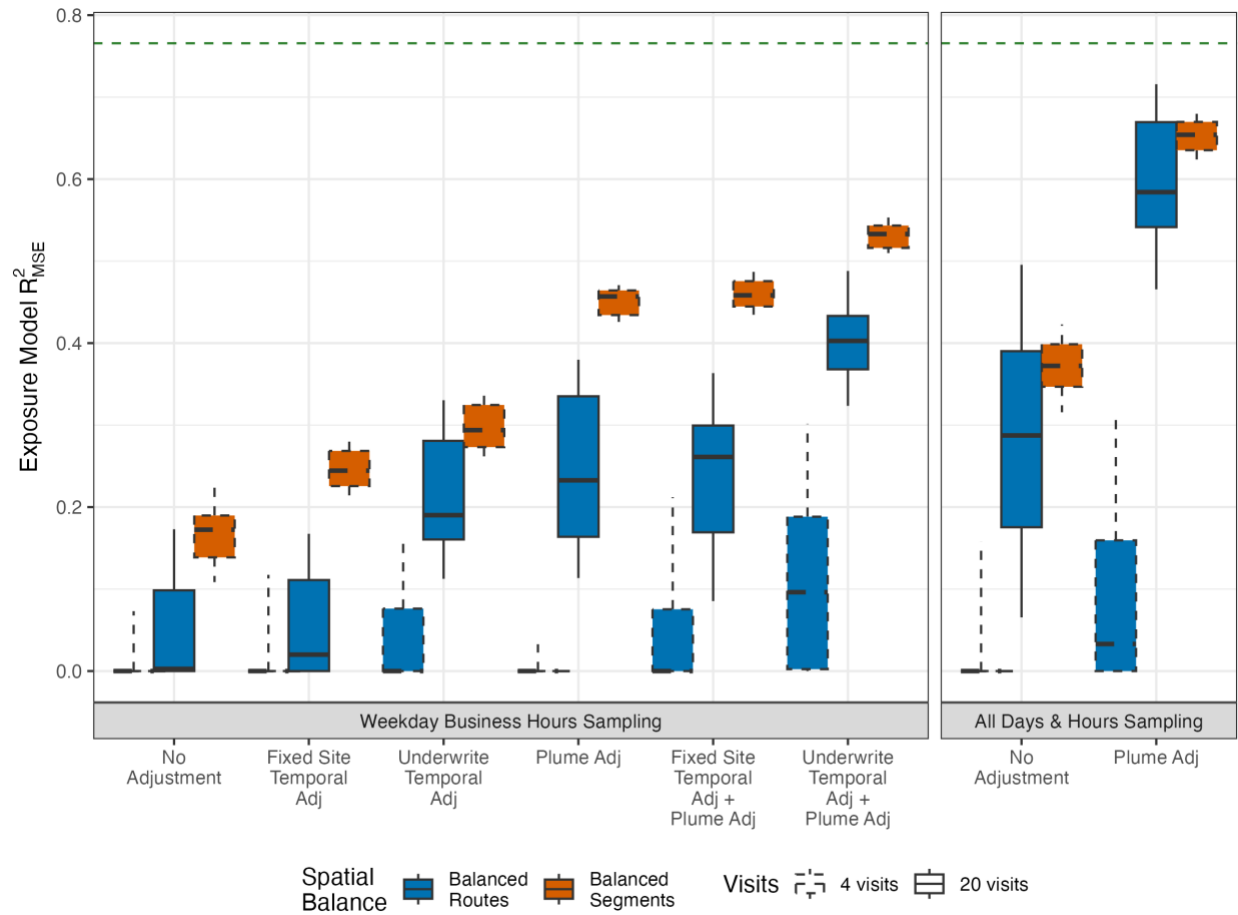


Figure S4. Out-of-sample PNC (pt/cm³) exposure model performances for on-road campaigns ($N=30$ campaigns per combination - i.e., boxplot) when sampling at the segment- vs. route-level. Compared to analyses that sample 4 visits per location at the segment level, analyses that consider logistical field constraints and sample entire routes generally perform worse, even when locations are visited 20 times (5 times as much). R^2_{MSE} is based on a comparison of the predicted PNC at 309 roadside locations to the annual average site estimates at those locations from stationary measures. The boxes illustrate the median and IQR, while the whiskers extend to the 10th and 90th percentiles. The dashed line shows the performance of the stationary roadside model.

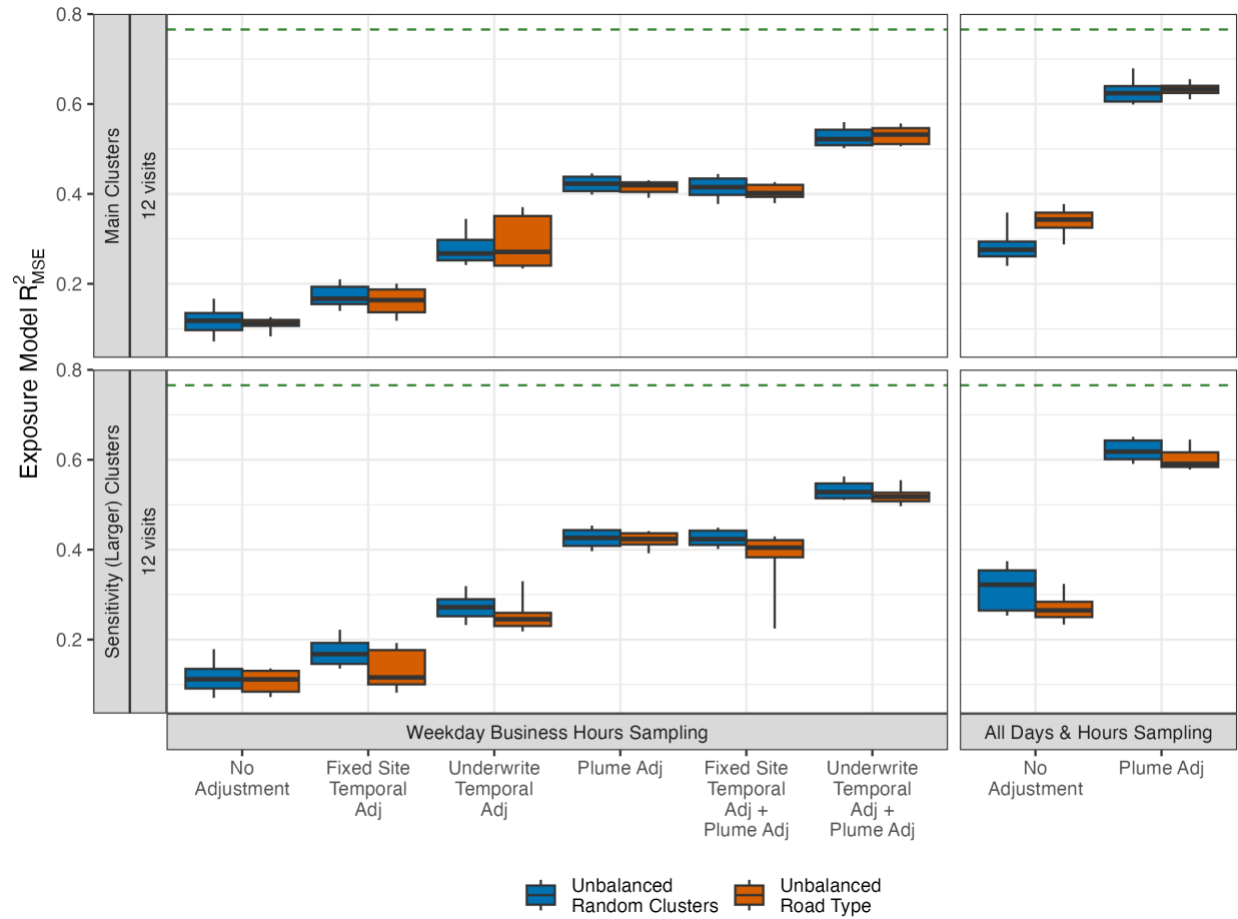


Figure S5. Out-of-sample PNC (pt/cm^3) exposure model performances for on-road campaigns ($N=30$ campaigns per combination - i.e., boxplot) based on spatial clusters used in the main analysis and alternative, larger clusters. Results are for campaigns with 12 visits per location. R^2_{MSE} is based on a comparison of the predicted PNC at 309 roadside locations to the annual average site estimates at those locations from stationary measures. The boxes illustrate the median and IQR, while the whiskers extend to the 10th and 90th percentiles. The dashed line shows the performance of the stationary roadside model.

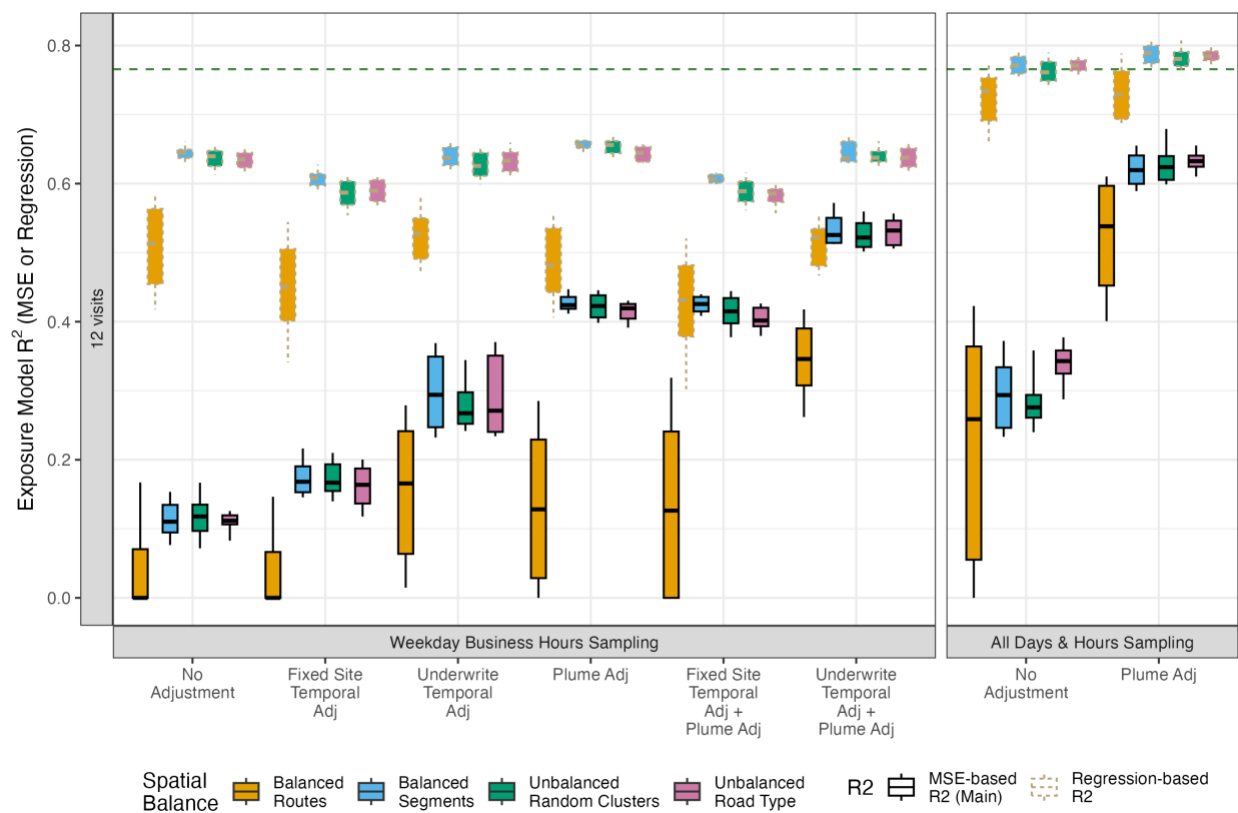


Figure S6. Out-of-sample PNC (pt/cm^3) exposure model performances for on-road campaigns ($N=30$ campaigns per combination - i.e., boxplot) evaluated using R^2_{reg} for 12-visit campaigns. R^2_{reg} is based on a comparison of the predicted PNC at 309 stationary locations and the annual average site estimate based on stationary measures. The boxes illustrate the median and IQR, while the whiskers extend to the 10th and 90th percentiles. The dashed line shows the performance of the stationary roadside model for respective R^2 , which is the same for R^2_{MSE} and R^2_{reg} .

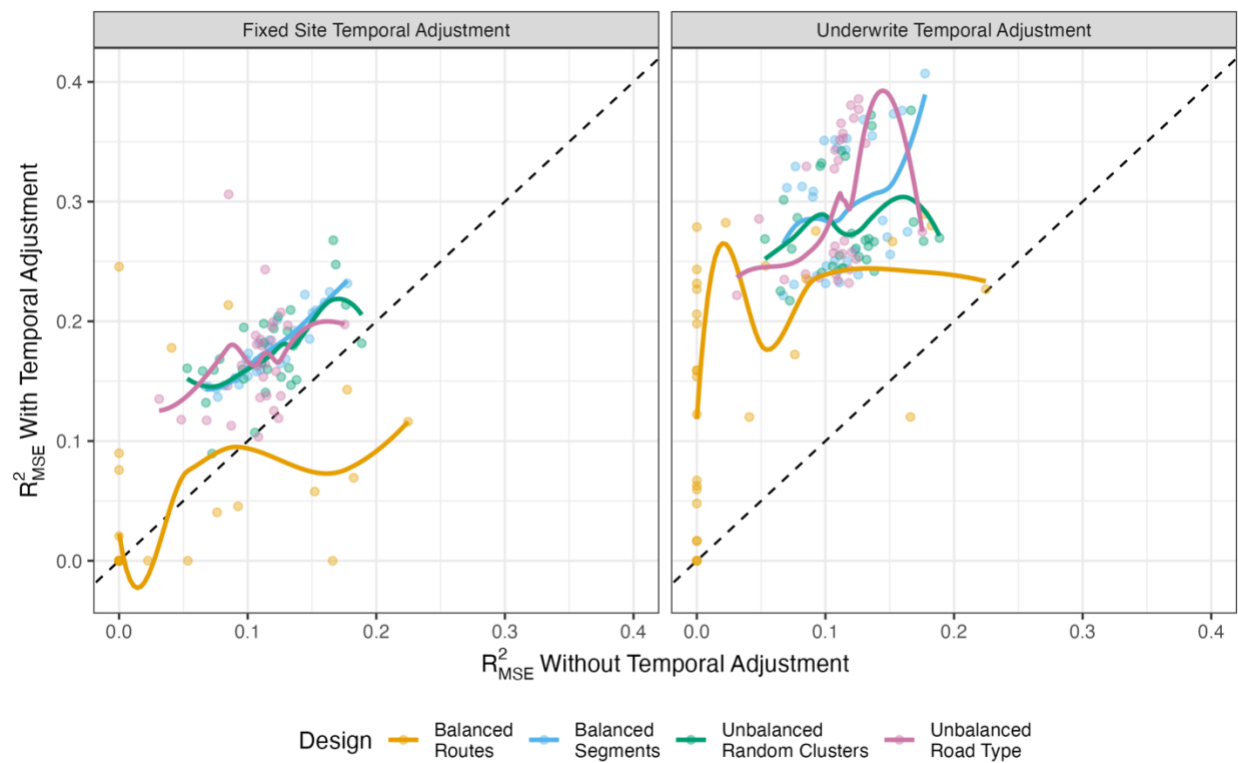


Figure S7. Comparison of R^2_{MSE} for a subset of 12-visit business hours sampling designs with and without temporal adjustment. The diagonal dashed line is the 1-1 line; values above this line indicate better resulting R^2_{MSE} values when compared to unadjusted campaigns.

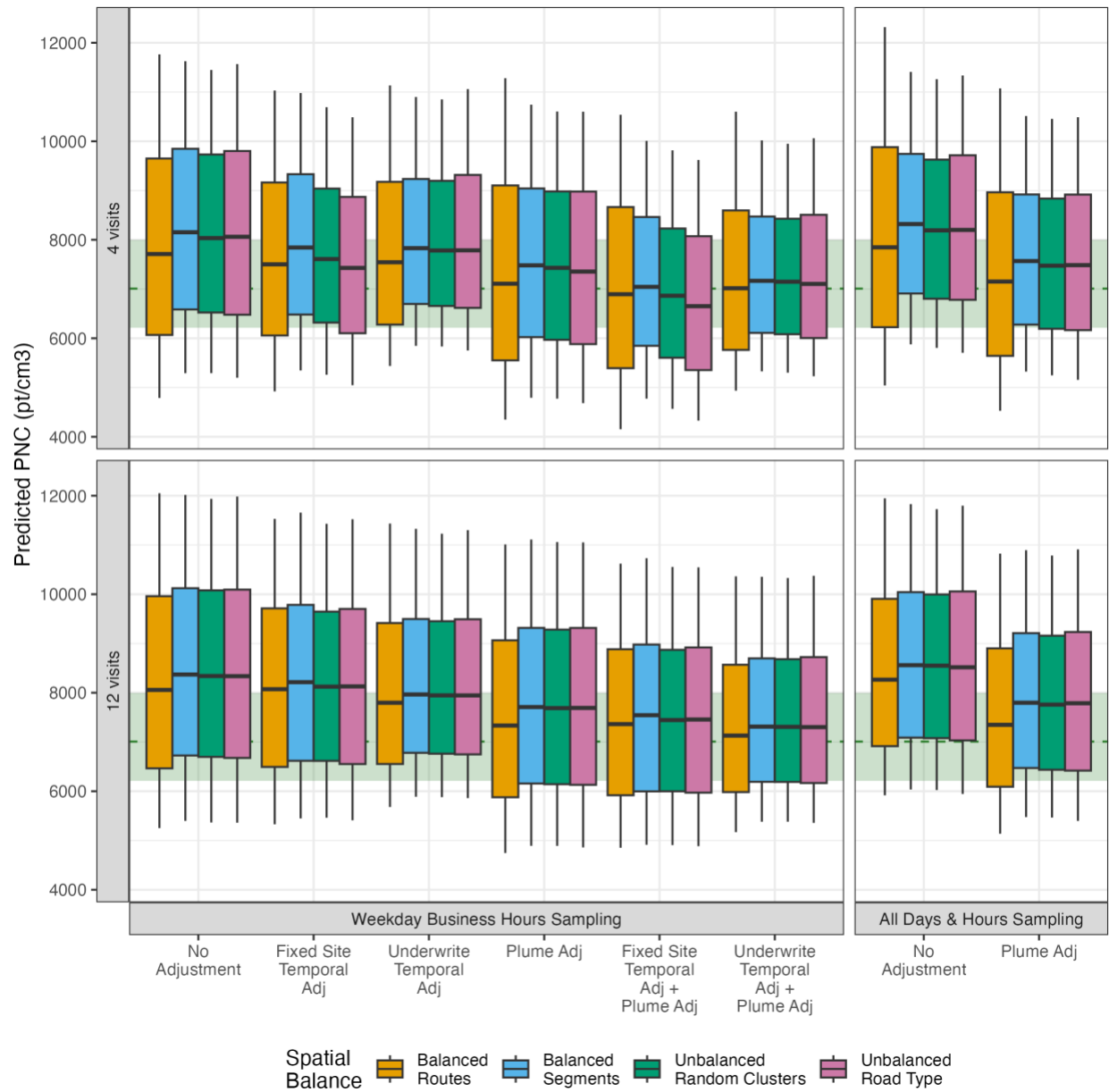


Figure S8. Predicted five-year average PNC exposure for ACT participants across 30 sampling campaigns for each design (i.e., boxplot). The green dashed line and area represent the median and IQR of cross-validated PNC predictions from the reference all-data stationary exposure model. The boxes illustrate the median and IQR, while the whiskers extend to the 10th and 90th percentiles.

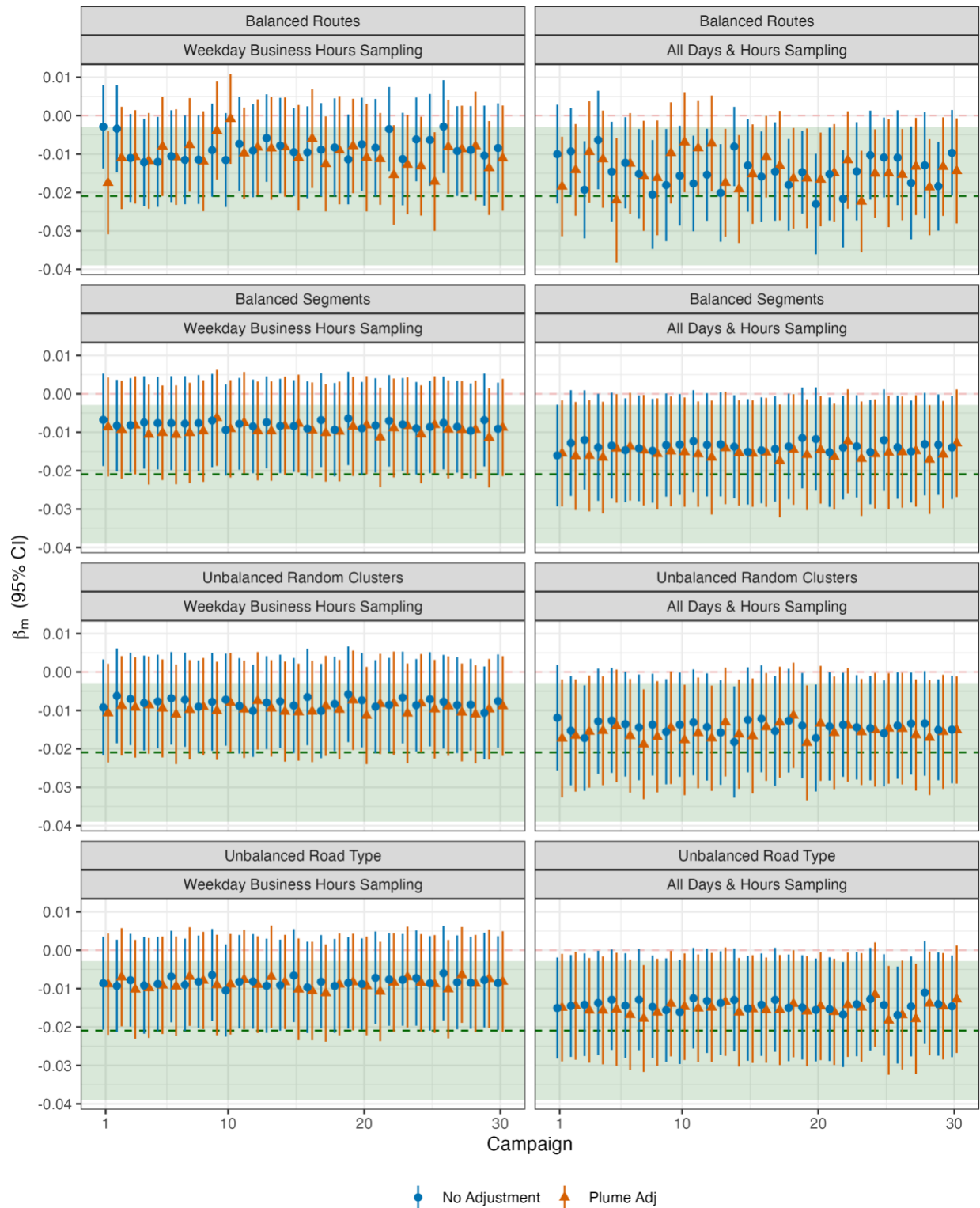


Figure S9. Point estimates and 95% confidence intervals (CI) from reduced health models of the adjusted association between PNC (1,900 pt/cm³) and CASI-IRT for selected campaigns. The green line and shaded area represent the point estimate and 95% CI from the roadside all-data exposure model, while the red line indicates no association.

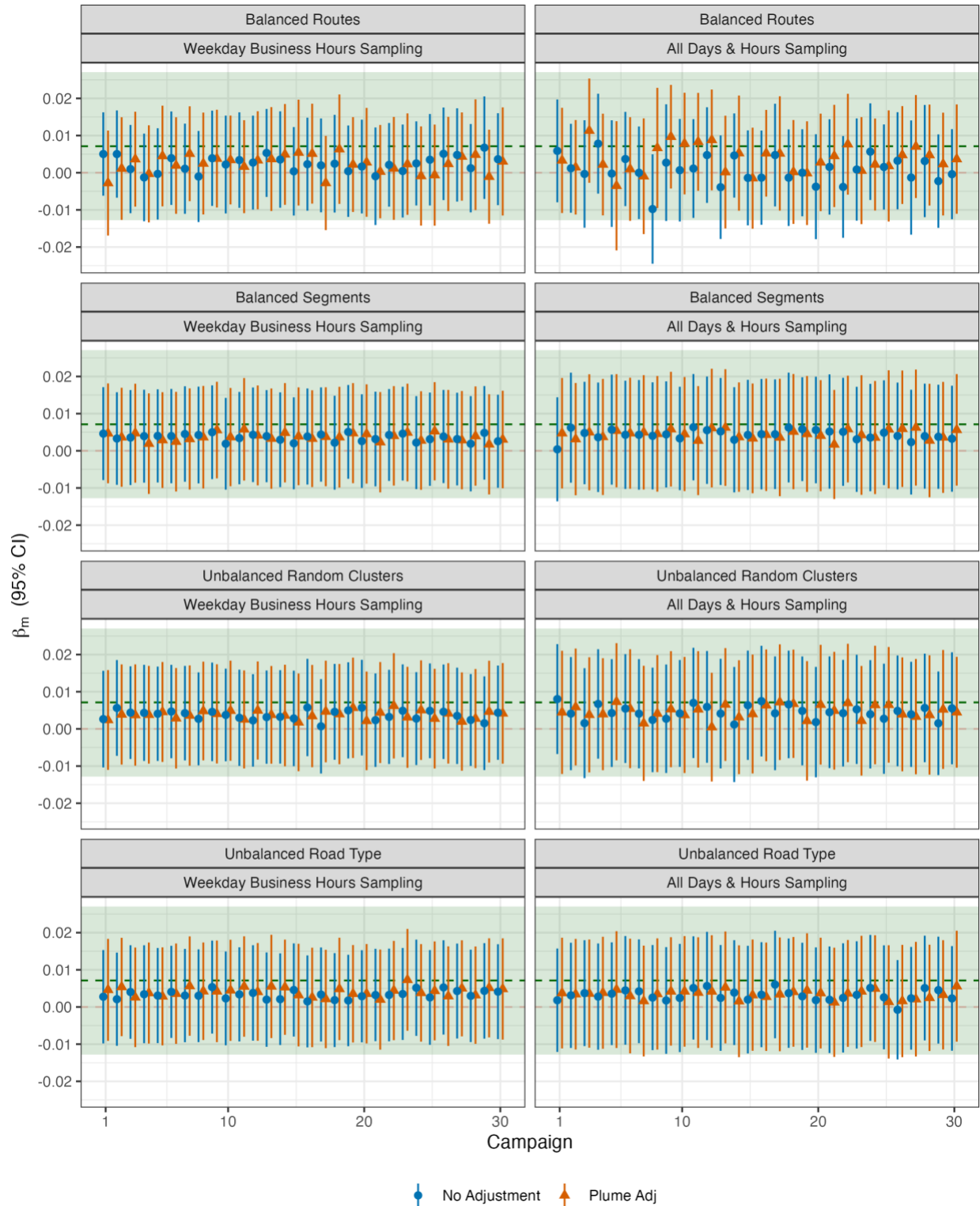


Figure S10. Point estimates and 95% confidence intervals (CI) from primary health models of the adjusted association between PNC ($1,900 \text{ pt/cm}^3$) and CASI-IRT for selected campaigns. The green line and shaded area represent the point estimate and 95% CI from the roadside all-data exposure model, while the red line indicates no association.

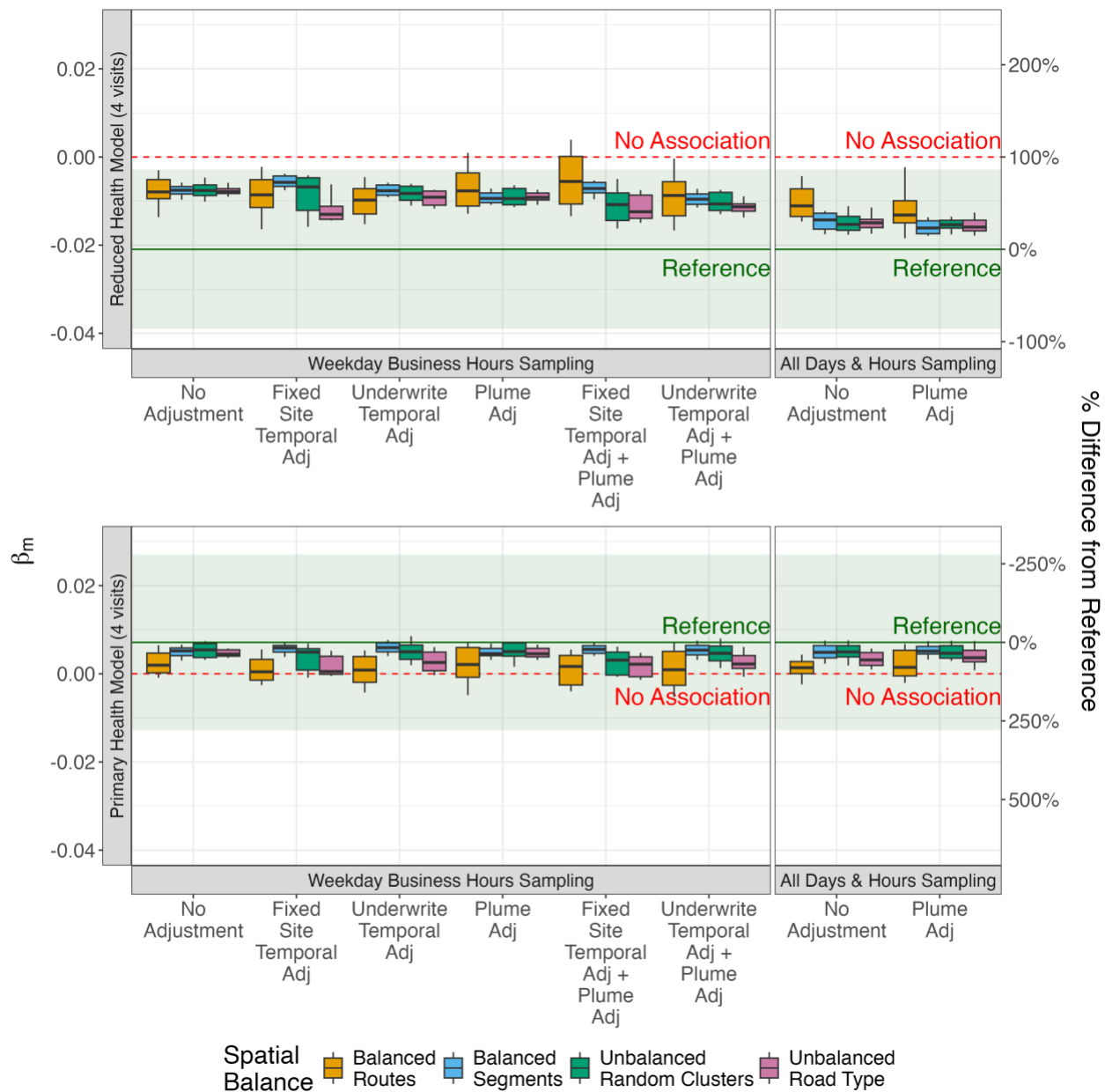


Figure S11. Estimated association between cognitive function (CASI-IRT) and PNC exposure (per 1,900 pt/cm³) for 4-visit campaigns. The dashed green line and colored area represent the reference health estimate and 95% CI from the roadside exposure model. The dashed red line indicates no association. Boxes show the median and IQR; whiskers illustrate the 10th and 90th percentiles of on-road campaign point estimates.