Research Report 226, *Comparison of Long-Term Air Pollution Exposure from Mobile and Routine Monitoring, Low-Cost Sensors, and Dispersion Models,* by G. Hoek et al.

INTRODUCTION

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

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

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

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

Dr. Hoek and colleagues proposed to compare the performance of a suite of air pollution exposure assessment methods in the Netherlands, including a comparison of health effects estimates among different methods. The HEI Research Committee recommended the study for funding because it would compare different exposure modeling approaches, including models based on data from low-cost sensors. They also appreciated the focus on UFPs and the leveraging of a wealth of data including three cohort studies.

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

SCIENTIFIC AND REGULATORY BACKGROUND

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

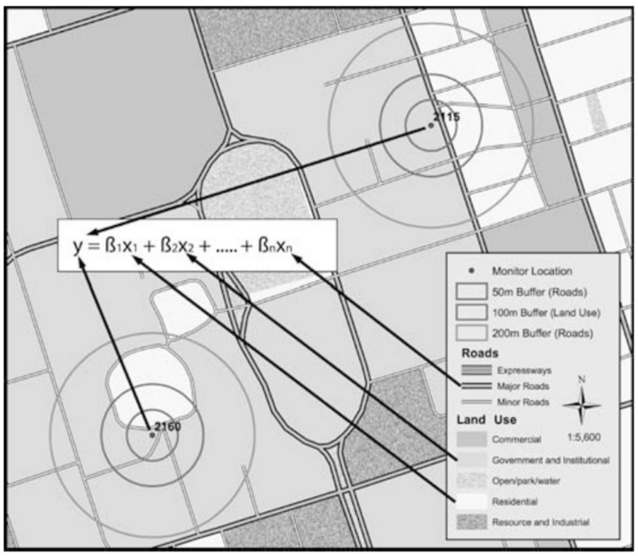
Epidemiological studies have used different approaches to address these challenges, including land use regression (LUR) models based on fixed-site routine monitoring, low-cost sensor networks, mobile monitoring, and dispersion models.

LUR models are empirically derived by regressing observations of air pollution at a number of locations on land use variables derived via geographic information systems (GIS). Typically, LUR models are developed using a variety of algorithms, such as supervised stepwise procedures or machine-learning algorithms, to identify which land use

Dr. Gerard Hoek's 3-year study, "Comparison of Long-term Air Pollution Exposure Assessment Based on Mobile Monitoring, Low-Cost Sensors, Dispersion Modelling, and Routine Monitoring-Based Models (CLAIRE)," began in June 2020. Total expenditures were $800,000. The draft Investigators' Report from Hoek and colleagues was received for review in February 2024. A revised report, received in July 2024, was accepted for publication in September 2024. During the review process, the HEI Improved Exposure Assessment Studies Review Panel and the investigators had the opportunity to exchange comments and clarify issues in the Investigators' Report and the Panel's Commentary. As a co-investigator of the Hoek Investigators' Report, Dr. Ulrike Gehring was not involved in its evaluation by the Panel.

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

variables are the best predictors of local concentrations of air pollution (Hoek 2017). Predictor variables often include indicators of land use, proximity to or density of nearby roads, traffic intensity, and population density. After fitting, the LUR model is used to estimate pollutant concentrations where no measurements are available in the study area, and ultimately, to predict exposures at locations important for epidemiological studies (**Commentary Figure** 1).



Commentary Figure 1. Illustration of elements of a LUR model (Jerrett et al. 2005).

LUR models can be based on fixed-site measurements, mobile measurements, or a combination of both. Mobile monitoring strategies can involve on-road measurements made while driving predefined routes, or repeated short-term measurements made while in a parked vehicle at various locations. LUR models can also be based on existing routine monitoring networks and specifically designed monitoring campaigns, either using research-grade measurement devices or low-cost sensors.

LUR models are empirical models in contrast to dispersion or chemical transport models, which are often deterministic models that are based on physical principles and estimated emissions. A wide variety of dispersion models exist that differ in their spatial scales (e.g., street, urban, regional, continental, or even global) and the processes they include (e.g., dispersion only or dispersion plus chemical transformations). The quality of the input data is a key determinant of the performance of any air pollution model (Hoek 2017). Recognizing the limitations of any single approach, researchers have also developed hybrid models that combine monitoring data, land use data, satellite observations, and dispersion models.

Little information is available about the relative performance of the various approaches to assess long-term exposure to traffic-related air pollution, and in turn, how those different approaches might influence health effects estimates. As noted earlier, important limitations and challenges remain when predicting long-term air pollution exposure for pollutants that vary highly in space and time.

Exposure models are applied in epidemiological studies that underpin the air quality standards and guidelines. Governments in the United States and Europe have recently moved toward more stringent $PM_{2.5}$ annual standards — 9 and 10 μg/m^3, respectively — which align more closely with the 2021 WHO Air Quality Guidelines of 5 μg/m^3. A more stringent NO_2 annual standard was also set in Europe (**Commentary Table 1**). There are no specific ambient air quality standards or guidelines for UFPs and BC, and regulatory agencies do not commonly measure them. Hence, international or national standard methods to characterize them have not been established (HEI 2010; HEI Review Panel on Ultrafine Particles 2013). Although no air quality guidelines have been developed for UFPs and BC, the WHO has provided "good practice statements" for these pollutants geared toward additional monitoring, mitigation, and epidemiological research (WHO 2021).

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

Organization	Annual $PM_{2.5}$ (μg/m^3)	Annual NO_2 (μg/m^3)	Notes
US EPA (2024)	9	100	NAAQS
US EPA (Previous)	12	100	Previous NAAQS
EU (2024)	10	20	Limit value for 2030
EU (Previous)	25	40	Previous limit value
WHO (2021)	5	10	Air Quality Guidelines
WHO (Previous)	10	40	Previous Air Quality Guidelines

NAAQS = National Ambient Air Quality Standards.

SPECIFIC AIMS

The specific aims of Dr. Hoek's study were to accomplish the following:

1. Develop long-term ambient air pollution exposure estimates for selected epidemiological studies based on low-cost sensors, mobile and fixed-site monitoring, and dispersion modeling
2. Compare different exposure assessment methods in terms of their ability to predict the spatial variation of long-term average concentrations using independent validation data
3. Compare different exposure assessment methods in terms of air pollution effect estimates in selected epidemiological studies

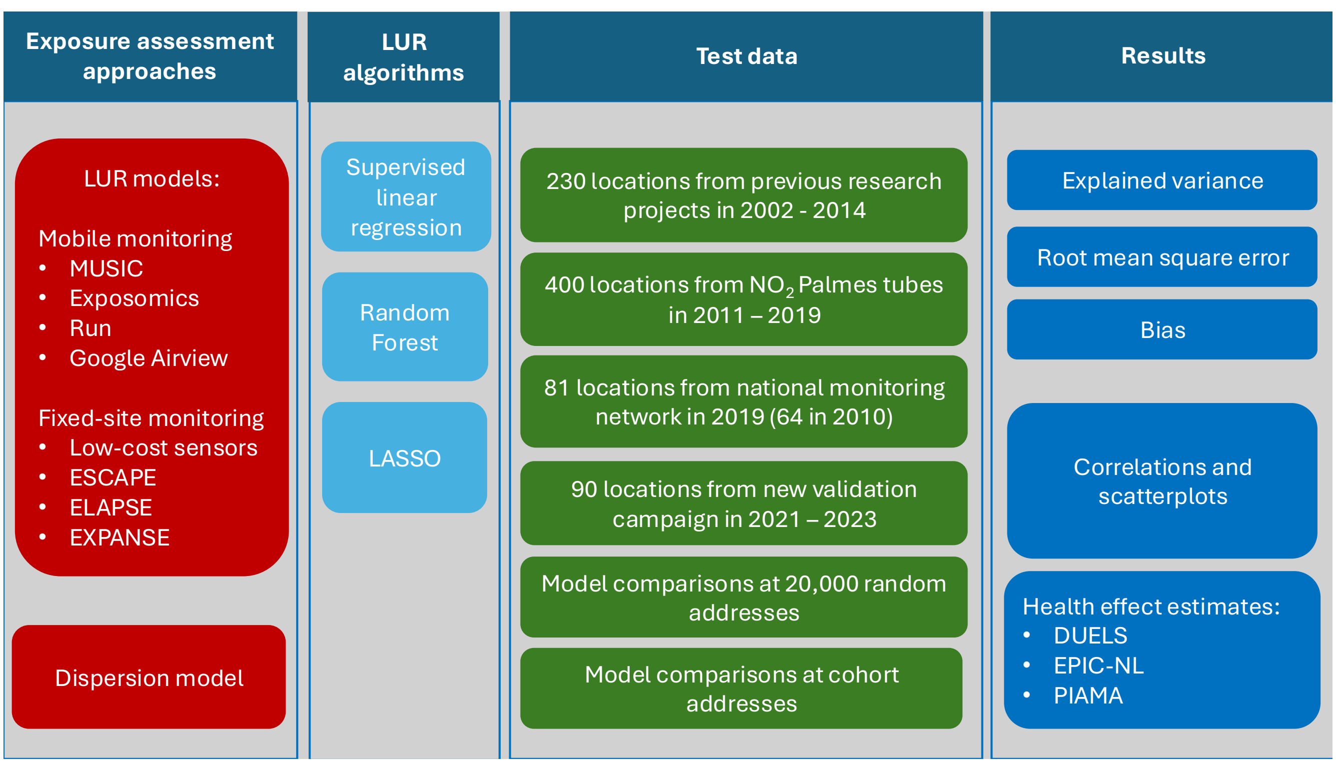
SUMMARY OF APPROACH AND METHODS

Dr. Hoek and colleagues compared the performance of a suite of air pollution exposure assessment methods in the Netherlands. The predictions of the exposure models were compared at 20,000 random residential addresses in the Netherlands and tested on existing and new validation data at residential locations where appropriate (**Commentary Figure 2**). The investigators also conducted epidemiological analyses in three cohort studies to compare health effect estimates of the different exposure assessment methods. They assessed four pollutants (UFPs, BC, $PM_{2.5}$, NO_2) and used existing data and models from previous collaborative projects where available.

EXPOSURE ASSESSMENT APPROACHES

Characteristics of the various exposure assessment approaches, including pollutants, monitoring approach, modeling approach, monitoring area, and year are provided in **Commentary Table 2**. The investigators evaluated annual average outdoor air pollution concentrations using a suite of exposure assessment methods that differ in their monitoring data (low-cost sensors, mobile monitoring, and fixed-site routine or fixed-site study-specific monitoring using research-grade instruments) and modeling approaches (LUR models or air pollution dispersion models). In addition, they tested various model development algorithms in the LUR-based models beyond linear regression, including machine learning methods.

The exposure assessment approaches also differed in the pollutants covered, the monitoring area, the time of day, and the study period. The greatest diversity of methods was available for BC and $PM_{2.5}$, while results for UFPs were only available from mobile monitoring models from different monitoring areas and years. For two approaches (the dispersion model and the EXPANSE model), the investigators were able to produce exposure predictions for multiple years to



Commentary Figure 2. Schematic overview of the study design.

Commentary Table 2. Characteristics of the Various Exposure Assessment Approaches

Model Name	Pollutants	Monitoring Approach	Monitoring Area	Modeling Approach	Year	Reference
MUSIC	UFPs, BC, $PM_{2.5}$	Mobile monitoring While parked: 3 × 30 minutes per year in 161 locations On-road: 2,964 road segments	Two cities: Amsterdam, Rotterdam	LUR	2013	Kerckhoffs et al. 2016
Exposomics	UFPs, BC, $PM_{2.5}$	Mobile monitoring While parked: 3 × 30 minutes per year in 240 locations On-road: 5,236 road segments	Three cities: Amsterdam, Maastricht, Rotterdam	LUR	2014–2015	Kerckhoffs et al. 2017
RUN	UFPs, BC, $PM_{2.5}$	Mobile monitoring While parked: 3 × 30 minutes per year in 400 locations On-road: 14,392 road segments	Netherlands	LUR	2016–2017	Kerckhoffs et al. 2021
Google Air View	UFPs, BC, $PM_{2.5}$, NO_2	Mobile monitoring On-road: 46,664 road segments, five to 10 repeats	One city: Amsterdam	LUR	2019–2020	Kerckhoffs et al. 2022
Low-cost sensors	$PM_{2.5}$, NO_2	Fixed-site monitoring 1 × 6 months in 84 locations	Netherlands	LUR and data-fusion model	2021–2023	New for the current study
ESCAPE	BC, $PM_{2.5}$, NO_2	Fixed-site monitoring 3 × 14-day average per year on 40–80 locations	Netherlands	LUR	2009–2010	Beelen et al. 2013; Eeftens et al. 2012
ELAPSE	BC, $PM_{2.5}$, NO_2	Fixed-site monitoring Annual average from 543–2,399 routine monitoring locations 436 sites from ESCAPE for BC	Western Europe	Hybrid LUR	2010	De Hoogh et al. 2018
EXPANSE	$PM_{2.5}$, NO_2	Fixed-site monitoring Annual average from 699–3,176 routine monitoring locations	Europe	Hybrid LUR	2010–2019	Shen et al. 2022
Dispersion model	EC, $PM_{2.5}$, NO_2	Not applicable	Netherlands	Deterministic	2010–2019	Velders et al. 2021

assess changes in spatial patterns between 2010 and 2019. All exposure assessment approaches were developed on data collected before the COVID-19 pandemic in 2020, except for the new low-cost sensor models.

New Low-Cost Sensor Campaign

A new low-cost sensor campaign was conducted at 84 residential outdoor locations across the Netherlands in 2021–2023 as part of a larger citizen-science project coordinated by the National Institute for Public Health and the Environment (Wesseling et al. 2019). Locations were either close to busy roads (~75%) or (urban) background locations (25%). Continuous measurements of NO_2 and $PM_{2.5}$ were collected using the Alphasense B43F sensor and the Sensirion SPS30 particle counter in two consecutive half-year measurement periods. Data from the national monitoring network were used to account for temporal differences between the two measurement periods. For NO_2, sensor calibration was performed before the first and second 6-month periods using six co-located sensors alongside reference measurements. Night-time NO_2 measurements were used to adjust differences between sensors on a daily basis to account for potential drift of the sensor within the 6-month period. For this night calibration, measurements at nearby reference stations were used to recalibrate the sensors at the participants' home addresses. The $PM_{2.5}$ sensors were not calibrated.

Low-cost sensor data were used primarily in data-fusion models, where these data were combined with dispersion model estimates and data from the routine national monitoring network to explore the added value of low-cost sensor data for NO_2. In addition, LUR models were developed using low-cost sensor data for both NO_2 and $PM_{2.5}$.

Algorithms Tested

For the various LUR-based models, the investigators tested three model development algorithms: supervised linear regression, Least Absolute Shrinkage and Selection Operator (LASSO), and Random Forest. Supervised linear regression is widely applied, which starts with univariate linear regression models for each potential predictor, and chooses the model with the highest adjusted R^2 as the starting point. Additional predictor variables were included in the model if they improved the adjusted R^2 of the previous model step, and only if the variable coefficient was in the expected direction of effect (e.g., positive for traffic intensity).

Other algorithms have increasingly been applied to alleviate some of the concerns related to standard linear regression. LASSO is a shrinkage algorithm that is used to weight the predictor coefficients more reliably when the predictors are highly correlated. Random Forest is a machine learning algorithm that can model nonlinearity and the potentially complex interactions among predictors.

The predictors available for inclusion in the algorithms tested in this study were based primarily on the ESCAPE model and included land use, traffic, and population density variables. For the European hybrid models, satellite observations and air pollution chemical transport model estimates were also available.

Existing and New Validation Data

The predictions of the different exposure approaches and algorithms were tested using existing and new validation data that were not used in model development.

Existing validation data were used from various research projects at more than 230 residential outdoor locations across the Netherlands covering more than 10 years. In addition, extensive monitoring was conducted using Palmes tubes for NO_2 across the Netherlands at 400 locations from 2011 to 2019. Validation data for the various pollutants were collected using a variety of monitoring approaches that also used different monitoring areas and study periods. NO_2 was the pollutant with the most validation data. Data were available for UFPs and BC at only 87 locations in Utrecht and Amsterdam, with relatively short sampling times (up to a week) per site from 2002 to 2004 and in 2014.

Data from the routine national monitoring network were also used. The number of locations differed per pollutant and varied across years. For 2019, the most data were available for NO_2 (81 sites), fewer for $PM_{2.5}$ (49 sites), and even fewer for BC (31 sites). No data were available for UFPs because they are not measured routinely in the Netherlands.

To increase the number of UFPs and BC measurements for validation purposes, Hoek and colleagues conducted a new campaign to acquire independent validation data from 2021 to 2023 — after most COVID-19 restrictions were lifted. Measurements of $PM_{2.5}$ and NO_2 were also collected for completeness. In total, measurements were conducted at 90 residential outdoor locations across the Netherlands. Similar to the new low-cost sensor campaign, traffic sites were oversampled. 1-week measurements were collected at the 90 locations using various research-grade instruments, including the DISCmini for UFPs; at 31 of those sites, an additional 1-week measurement was performed in a different season. Measurements were not taken simultaneously at all sites because of equipment limitations; only four locations were measured simultaneously. Data from a reference site in the middle of the country were used to account for temporal differences between measurement weeks.

To represent long-term exposure, validation data were aggregated to an average concentration per location, which also accounted for temporal differences between measurement weeks where applicable.

Assignment of Exposure Estimates to Locations

The investigators assigned the annual air pollution model predictions to 20,000 random residential addresses, addresses of cohort participants, and validation locations using a standard Dutch database of geocoded addresses from 2019. For the assignment to the routine network sites, they used the nearest residential address; for all other validation data, the exact residential address was used because monitoring took place at residential home locations.

For all cohort participants, exposure data were assigned to the residential address at the time of recruitment to the cohort. For PIAMA, exposure was also assigned at the current address.

HEALTH STUDIES

Epidemiological analyses in three cohort studies were conducted to compare the health effect estimates of the different exposure assessment approaches and algorithms.

Study Population and Health Outcomes

For the health analyses, Hoek and colleagues selected three population-based cohorts in the Netherlands that differ in size, population, location, health outcomes, and study period (**Commentary Table 3**). DUELS is a very large administrative cohort that includes all Dutch adult citizens aged 30 years or older (10.8 million), starting from 2013 and following until 2019. DUELS was formed by linking census data, population registries, and death registries, but contains less detailed covariate data than the other selected cohorts. In addition, the investigators selected two smaller, "conventional" epidemiological cohorts with detailed information available on lifestyle factors (EPIC-NL and PIAMA). EPIC-NL is an adult cohort of mostly women (76%), whereas PIAMA is a birth cohort that has 20 or more years of follow-up data. Both DUELS and EPIC-NL were also included in the HEI-funded ELAPSE study that investigated the health effects of low levels of air pollution (Brunekreef et al. 2021).

DATA ANALYSES

Data Analysis Comparing Exposure Predictions and Performance

Hoek and colleagues compared annual average predictions of the different exposure assessment methods by preparing scatterplots and calculating Pearson correlations between predictions from the different models. They also prepared Bland-Altman plots to identify systematic differences between the exposure predictions. They calculated the correlations across years where available and across pollutants within methods. They assessed model performance in terms of explained variance (R^2), root-mean-square error (RMSE), and bias (model minus measurement) by comparing model predictions with validation data. Comparisons were made for three spatial domains: country-wide, four major cities (Amsterdam, Rotterdam, Utrecht, and The Hague), and Amsterdam.

Data Analysis in the Health Studies

The investigators applied Cox proportional hazards models to assess the association between air pollution exposure from the different exposure assessment approaches and various health outcomes in DUELS and EPIC-NL. For PIAMA, they applied linear regression and discrete-time hazard models to investigate the associations between the different air pollution exposure estimates and lung function and asthma incidence, respectively.

Commentary Table 3. Key Study Characteristics of the Three Health Studies in the Netherlands

Study Name	Study Population	Location	Health Outcomes	Study Period	Sample Size (rounded)	Age	Reference
DUELS	Adult administrative cohort	Nationwide	Natural-cause and cause-specific mortality	2013–2019	10.8 million	30+	Brunekreef et al. 2021; Klompmaker et al. 2021
EPIC-NL	Adult cohort	Four cities: Amsterdam, Doetinchem, Utrecht and Maastricht	Natural-cause mortality Coronary events Stroke events	1993–2013	34,000	20+	Brunekreef et al. 2021
PIAMA	Birth cohort	3 regions: northern, middle, and southwestern	Asthma incidence	1996–2016	3,700	0–20	Yu et al. 2022
		3 regions: north, middle, and southwest	Lung function (FEV_1, FVC)	1996–2012	700	0–16	Yu et al. 2021

DUELS = Dutch Environmental Longitudinal Study; EPIC-NL = European Prospective Investigation into Cancer and Nutrition: Netherlands cohort; FEV_1 = forced expiratory volume in 1 second; FVC = forced vital capacity; PIAMA = Prevention and Incidence of Asthma and Mite Allergy.

The main results for DUELS were adjusted for age, sex, and individual and area-level socioeconomic status. In addition, results were adjusted for various lifestyle factors, such as smoking and diet for EPIC-NL and PIAMA.

The main analyses focused on single-pollutant models, although limited two-pollutant models were applied for DUELS for natural-cause mortality and a subset of exposure approaches. The investigators expressed the effect estimates using fixed increments across methods (5 µg/m³ for $PM_{2.5}$, 10 µg/m³ for NO_2, 1 µg/m³ for BC, and 5,000 p/cm³ for UFPs) and method-specific interquartile ranges of exposure.

To further explore if and why effect estimates differed among exposure approaches, the investigators conducted some additional analyses for DUELS and PIAMA. For example, they used scatterplots and linear regression analysis to evaluate whether the performance of the different exposure models and the predicted exposure contrast were associated with the effect estimate. Furthermore, they conducted a random-effects meta-analysis with the different exposure assessment methods to illustrate the degree of heterogeneity potentially introduced in meta-analyses, because in meta-analysis it is common practice to combine the results of studies using different exposure assessment methods.

SUMMARY OF RESULTS

COMPARISON OF EXPOSURES ACROSS DIFFERENT METHODS

In most comparisons, the investigators observed only small differences in performance (<0.1 difference in R^2) among the three LUR algorithms (supervised linear regression, Random Forest, or LASSO). Hence, this summary focuses on the results of the supervised linear regression models.

In some cases, the predicted exposure levels and exposure contrasts varied widely between methods. For example, the interquartile range at the 20,000 residential locations differed by a factor of up to 14 for BC across methods and was even higher for $PM_{2.5}$ (a factor of 17). Exposure contrast factors of 4 and 8 were reported for UFPs and NO_2, respectively. In general, the mobile monitoring models predicted higher BC and $PM_{2.5}$ levels and contrasts at residential sites than other models, probably related to the on-road monitoring. Exposure contrasts for $PM_{2.5}$ were comparatively low for all models except the mobile monitoring models (the interquartile range varied from 0.9 to 2.3 µg/m³ in all nonmobile monitoring models; a factor of 3).

The various exposure models generally resulted in moderately to highly correlated exposure predictions for BC, NO_2, and UFPs at residential sites across the Netherlands (correlation coefficient $R > 0.7$). In contrast, the correlations for $PM_{2.5}$ between the different models were lower ($R < 0.4$), particularly for the poorly performing mobile monitoring models.

Exposure predictions for different years in the period of 2010–2019 (dispersion model and EXPANSE model) were highly correlated for BC, NO_2, and $PM_{2.5}$ ($R > 0.9$), and indicated stable spatial contrasts over a 10-year period. Exposure levels of BC, NO_2, and $PM_{2.5}$ declined during this period.

Exposure models explained a low to moderate amount of spatial variance in UFPs and NO_2 at the new validation sites (R^2 between 16% and 47%), and a positive bias was reported for all models (models overestimated measurements up to 13,000 p/cm³ for UFPs and 17 µg/m³ for NO_2).

Most models explained a moderate amount of spatial variance in past exposures at existing validation sites for BC and NO_2 ($R^2 > 0.5$) and explained less variance for UFPs (R^2 >0.25). Most models predicted past exposures of $PM_{2.5}$ poorly (**Commentary Table 4**).

Commentary Table 4. The Explained Variance (in %) of the Mobile, Fixed-Site, and Air Pollution Dispersion Exposure Assessment Approaches at Existing Validation Sites

	2019 National Monitoring Network			2019 Palmes Tubes, National	2007 TRACHEA, National	2002–2004 RUPIOH, Amsterdam	2014 Exposomics, Amsterdam and Utrecht
Exposure Assessment Approaches	**BC ($N = 31$)**	**$PM_{2.5}$ ($N = 49$)**	**NO_2 ($N = 81$)**	**NO_2 ($N = 404$)**	**NO_2 ($N = 144$)**	**UFPs ($N = 46$)**	**UFPs ($N = 41$)**
Mobile monitoring[a,b]	45	1	69	61	73	24	46
Fixed-site monitoring[a]	72	5	63	52	62	NA	NA
Air pollution dispersion	61	18	75	62	77	NA	NA

NA = not available; RUPIOH = Relationship between Ultrafine and fine Particulate matter in Indoor and Outdoor air and respiratory Health; TRACHEA = Traffic-Related Air pollution and Children's respiratory HEalth and Allergies.
[a] Based on LUR models using supervised linear regression.
[b] Includes on-road and short-term parked monitoring.

The addition of low-cost sensor data did not improve NO_2 estimates in models that combined dispersion model estimates and data from the routine national monitoring network.

COMPARISON OF HEALTH EFFECTS FOR THE DIFFERENT METHODS

In most cases, the application of the various exposure assessment approaches in the health studies led to similar findings in terms of direction (positive, null, or negative associations) between air pollutant exposure and various mortality and morbidity outcomes (**Commentary Table 5**). There were no consistent differences in effect estimates between exposure assessment approaches based on mobile monitoring, fixed-site monitoring, or dispersion models. These results were consistent with the typically moderate-to-high correlations of the different exposure assessment approaches. Positive (adverse) associations were observed most clearly in DUELS and to a lesser extent in PIAMA. Null or sometimes even negative associations were reported in the EPIC-NL cohort.

The strength (magnitude) of the positive associations, however, differed based on the exposure assessment method. For example, the association for natural-cause mortality in DUELS for a 1 μg/m³ increase in BC ranged from a hazard ratio of 1.007 (95% confidence interval: 1.005–1.009) to 1.07 (1.05–1.09). Typically, differences in the magnitude of the associations across methods were smaller when expressed per method-specific interquartile range. Heterogeneity in the effect estimates was high (I^2 > 88%) for natural-cause mortality in DUELS, but low to moderate for the effect estimates reported in PIAMA.

Factors that explained some of the heterogeneity of effect estimates included the performance of the model at validation sites and the predicted exposure contrast. The year of the exposure model did not explain the heterogeneity.

Commentary Table 5. Summary of Positive, Null (Nonstatistically Significant), or Negative Findings for Selected Health Outcomes in the Three Cohort Studies

	DUELS				EPIC–NL				PIAMA			
	Natural-Cause Mortality								Asthma Incidence			
Model Name[a]	**UFPs**	**BC**	**NO_2**	**$PM_{2.5}$**	**UFPs**	**BC**	**NO_2**	**$PM_{2.5}$**	**UFPs**	**BC**	**NO_2**	**$PM_{2.5}$**
MUSIC (mobile)	0	+	NA	+	0	0	NA	–	+	+	NA	+
MUSIC (short-term parked)	+	+	NA	+	0	–	NA	0	0	+	NA	0
Exposomics (mobile)	+	+	NA	+	–	–	NA	0	+	0	NA	+
Exposomics (short–term parked)	+	+	NA	+	0	0	NA	0	0	0	NA	+
RUN (mobile)	+	+	NA	+	0	–	NA	0	0	0	NA	0
RUN (short-term parked)	+	+	NA	+	0	0	NA	0	+	0	NA	+
Google Air View	+	+	+	NA	0	0	+	NA	0	+	+	NA
Low-cost sensors	NA	NA	0	+	NA	NA	0	+	NA	NA	0	0
ESCAPE	NA	+	+	0	NA	0	0	0	NA	+	+	+
ELAPSE (2010)	NA	+	+	0	NA	0	0	0	NA	+	+	+
EXPANSE (2010)	NA	NA	+	+	NA	NA	0	+	NA	NA	+	+
EXPANSE (2013)	NA	NA	+	+	NA	NA	0	0	NA	NA	+	+
EXPANSE (2016, 2019)	NA	NA	+	+	NA	NA	0	+	NA	NA	+	+
Dispersion model (2010, 2013, 2016, 2019)	NA	NA	+	+	NA	NA	0	0	NA	NA	+	+

NA = not available.
[a] Models from supervised linear regression were chosen for the LUR models.

HEI IMPROVED EXPOSURE ASSESSMENT STUDIES REVIEW PANEL'S EVALUATION

In its independent review of the study, the Panel thought the study was well-motivated and effectively leveraged a wealth of air pollution and health data. They thought the study was comprehensive with thorough analysis and findings that will be of broad interest and value to a wide audience. The study documented that the various exposure models generally resulted in moderately to highly correlated exposure predictions for all pollutants except $PM_{2.5}$. Findings on the presence of an association with various mortality and morbidity outcomes were similar, albeit with sometimes notable differences in the magnitude of the associations.

STRENGTHS OF THE STUDY

The Panel noted several strengths of the research. First, the comparison of a large suite of exposure models commonly used in epidemiological studies was notable and made the results relevant and widely applicable. The study included a variety of models such as LUR models based on fixed-site routine monitoring, mobile monitoring, and dispersion models. The study also investigated state-of-the-art hybrid models that combined monitoring data, land-use data, satellite observations, and estimates from chemical transport models. Moreover, the study assessed four pollutants, including UFPs and BC for which the evidence base is still limited, partly due to the lack of comprehensive routine monitoring.

Second, the Panel thought the specific assessment of the potential role of low-cost sensors in developing exposure models was particularly useful because low-cost sensors are increasingly used. The investigators performed state-of-the-art calibrations of the NO_2 low-cost sensors, yet they still did not find them helpful in predicting long-term exposures once other more common sources of exposure were included in prediction models.

Third, the extensive validation efforts at both the nationwide and city-specific domains and the reporting of multiple measures to test the performance of the exposure models are strengths. The investigators made use of new and existing validation data spanning a 20-year period. The comparisons of predictions at 20,000 randomly selected addresses and the cohort participants' addresses were useful for the evaluation. Moreover, the investigators reported multiple measures of performance — not only correlations and explained variance but also possible bias and systematic differences using, for example, Bland-Altman plots — thereby providing an in-depth performance assessment.

Fourth, the application of the various exposure models in relation to health outcomes was another strength. The investigators applied the various exposure estimates to three population-based cohorts in the Netherlands that differ in size, population, location, health outcomes, and study period. In particular, the health analysis for a very large population (10.8 million) that included all Dutch adult citizens aged 30 years or older was considered informative. The inclusion of both mortality and morbidity outcomes and the availability of lifestyle factors (e.g., smoking and diet) in the two smaller cohorts were also noted as strengths.

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

LIMITATIONS IN THE VALIDATION DATA

The Panel was impressed by the extensive validation efforts and the leverage of many datasets but noted some limitations, particularly in the utility of the new validation data. The short duration of the measurements, the nonsimultaneous sampling, and the small number of repeated measurements for the different seasons hampered the utility of these new data for validating the different exposure assessment approaches. The monitoring campaign had to be postponed because of the COVID-19 pandemic, and only limited equipment (4 instruments) was available. Eventually, in 31 locations, 1-week measurements were available in two different seasons, but this monitoring approach fell short of the original plan to conduct measurements in 100 locations in two different seasons. Those issues limited the ability to extrapolate the measurements reliably to annual mean exposures and then compare them with the various predictions from annual average air pollution models.

Indeed, large temporal variation was evident for all four pollutants in the new measurement data, even after accounting for temporal differences between measurement weeks using a reference site. Moreover, the investigators did not interpret the data for BC and $PM_{2.5}$ because of the low correlation between the first and second week of measurements. The investigators added a post-hoc analysis where instead of comparing individual site average measurements and modeling, they ranked observations and averaged over 5 observations to address the temporality issue better, which indeed increased model performance. Regardless, the Panel concurred with the investigators that the temporal coverage of the new validation data was a limitation, but thought there was value for other researchers in the lessons learned from this effort, as nicely described in the report's discussion section.

The comparisons using existing RUPIOH and Exposomics validation data were limited by similar temporal coverage issues. The cleanest comparison set was from TRACHEA, where simultaneous 1-week measurements of NO_2 were available in four seasons in 2007 at 144 locations. The comparison using fixed-site national monitor data with complete temporal coverage was comprehensive, but problematic for a few models (i.e., dispersion models or low-cost sensor models) because these models were developed using those same data. Moreover, to compare the dispersion model results using fixed-site national monitor data, the investigators used the nearest residential address instead of the exact monitoring site location, which could have influenced the results.

Due to limitations in the available validation data, and because model results were generally highly correlated, no clear preference for a specific exposure assessment approach or model development algorithm was stated. Hence, more work is needed to evaluate alternative exposure models and assessment approaches. Despite those limitations, the validation efforts were extensive and thorough and should interest many readers.

FOCUS ON ESTIMATING OUTDOOR RESIDENTIAL EXPOSURE

Like many other outdoor air pollution and health studies, the current study only focused on estimating outdoor concentrations at residential locations. There is also interest in how this outdoor pollution contributes to personal exposure, although there is no reason to expect this to vary by the exposure assessment methods evaluated in the current study. Steps toward assessing personal exposure to outdoor pollution include accounting for individual time–activity patterns, including commuting and time spent at nonresidential locations, rather than only residential addresses. Outdoor and indoor concentration measurements can also be paired to estimate infiltration factors. These steps are challenging, costly, and demanding to perform for large cohorts, and thus were understandably beyond the scope of this study.

Lack of consideration of infiltration rates and time–activity data adds exposure measurement error, which is often assumed to bias the estimated outdoor air pollution and health estimates toward the null, although the nature of the potential bias cannot be fully known (Kioumourtzoglou et al. 2014; Sheppard et al. 2012). A recent review, which was part of the upcoming HEI report of de Hoogh funded under the same RFA as the current study, reported similar health effects among studies using residential vs. time-integrated exposure assessment of air pollution exposure, documenting that the overall bias in epidemiological studies might likely be small (Hoek et al. 2024).

Methodological complexity and lack of data prevented the investigators from propagating exposure measurement error into the health effects estimations. How to propagate exposure measurement error into health effects estimation in long-term air pollution and health studies remains an area of future research (Samoli and Butland 2017).

HETEROGENEITY IN EFFECT ESTIMATES

Substantial heterogeneity was found in the magnitude — but not direction — of the air pollution associations with the different exposure assessment methods within and across studies. The Panel appreciated the investigators' attempts to explore heterogeneity, even though it was not fully explained by the various factors considered in the study. Several factors were considered, such as the performance of the model, the predicted exposure contrast, and the year of the exposure model. Some of the factors, however, such as spatial resolution, temporal coverage, number of sites, and modeling domain, are difficult to separate from other factors because many of them are related, as noted by the authors. The Panel commented that the heterogeneity could be informative and warrants further examination. Relatedly, the Panel thought the discussion regarding the null or sometimes even negative associations in the EPIC-NL cohort could have been deepened.

The Panel found the implications from an illustrative meta-analysis of the different exposure assessment methods intriguing for evidence synthesis. The investigators showed elegantly that differences in exposure assessment can already lead to high heterogeneity in the same epidemiological study. Hence, it is unsurprising that high heterogeneity is typically reported in systematic reviews and meta-analyses of the health effects of air pollution (e.g., Chen and Hoek 2020; Huangfu and Atkinson 2020; Kasdagli et al. 2024; Orellano et al. 2024) because these analyses compare and integrate results from studies that differ in many methods and design features beyond the exposure assessment. Similarly, in HEI's low-exposure epidemiology initiative, substantial heterogeneity was found in the magnitudes of the positive associations within and across studies (Boogaard et al. 2024), with heterogeneity only slightly reduced in a harmonized analysis using the same exposure model, outcome definition, population age, covariates, and statistical models (Chen 2023). The meta-analysis in the current study implies that for evidence synthesis, caution is warranted in downgrading a body of literature based on heterogeneity statistics — an important point also made in a paper that summarizes the lessons learned from HEI's systematic review on traffic-related air pollution (Boogaard et al. 2023; HEI 2022). This finding is important because heterogeneity in study results is often misinterpreted as a sign of weak or inconsistent evidence, potentially undermining confidence in a body of literature. Recognizing this nuance allows for a more accurate and fair synthesis of evidence, supporting well-informed public health and policymaking decisions.

Of note in the cohort applications is that the main analyses focused on single-pollutant models. This choice is understandable given the volume of results already generated by the project. However, it is important to note that conclusions drawn from two-pollutant models might differ importantly from those drawn from single-pollutant models. This was demonstrated by the limited two-pollutant models that were run using DUELS. In most analyses using the air pollution dispersion model, only one of the two pollutants remained statistically significantly associated with natural-cause mortality. In contrast, in analyses with the ELAPSE and EXPANSE models, both pollutants remained associated with that outcome. Future research should prioritize the inclusion of multipollutant models to enhance the robustness and accuracy of findings regarding the health effects of air pollution.

SUMMARY AND CONCLUSION

The study by Dr. Hoek and colleagues compared the performance of a suite of air pollution exposure assessment methods in the Netherlands for four pollutants (UFPs, BC, $PM_{2.5}$, NO_2). It included a comparison of health effects estimates derived from these methods. The predictions of the exposure models were compared at random residential addresses in the Netherlands and tested on existing and new validation data. Epidemiological analyses in three cohort studies were conducted to compare health effect estimates of the various exposure assessment methods. For all cohort participants, the various exposure estimates were assigned to the residential address at recruitment to the cohort.

The study leveraged a wealth of air pollution and health data from previous collaborative projects and included a large variety of models commonly applied in epidemiological studies, making the results relevant and widely applicable. The extensive validation efforts at both the nationwide and city-specific domains and the reporting of multiple measures to test the performance of the exposure models were considered additional strengths. The investigators used new and existing validation data spanning a 20-year period. Applying those models in relation to various health outcomes in three different cohorts was another strength. In particular, the health analysis for a very large population (10.8 million) that included all Dutch adult citizens aged 30 years or older was considered informative.

The study documented that the various exposure models generally resulted in moderately to highly correlated exposure predictions for all pollutants except $PM_{2.5}$. Findings on the presence of an association with various mortality and morbidity outcomes were similar, albeit with sometimes notable differences in the magnitude of the associations.

Although the Panel broadly agreed with the investigators' conclusions, some limitations should be considered when interpreting the results. The comparison of the different exposure approaches using new validation data was hampered by the short duration of the measurements, the nonsimultaneous sampling, and the small number of repeated measurements for the different seasons. Those issues limited the ability to extrapolate the measurements reliably to annual mean exposures and then compare them with the various annual average air pollution models. Some comparisons using existing validation data were limited by similar temporal coverage issues. Substantial heterogeneity was found in the magnitude — but not direction — of the air pollution associations of the different exposure assessment methods within and across studies. The Panel thought the heterogeneity was not fully explained by the various factors considered in the study and warrants further examination.

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

ACKNOWLEDGMENTS

The HEI Review Committee is grateful to the Improved Exposure Assessment Studies Review Panel for their thorough review of the study. The Committee is also grateful to Dan Crouse for his oversight of the study, to Hanna Boogaard for assistance with reviewing the report and in preparing its Commentary, to Tom Zaczkiewicz for editing this Report and its Commentary, and to Kristin Eckles and Hope Green for their roles in preparing this Research Report for publication.

REFERENCES

Beelen R, Hoek G, Vienneau D, Eeftens M, Dimakopoulou K, Pedeli X. 2013. Development of NO_2 and NO_x land use regression models for estimating air pollution exposure in 36 study areas in Europe: the ESCAPE project. Atmos Environ 72:10–23, *https://doi.org/10.1016/j.atmosenv.2013.02.037*.

Boogaard H, Atkinson RW, Brook JR, Chang HH, Hoek G, Hoffmann B, et al. 2023. Evidence synthesis of observational studies in environmental health: lessons learned from a systematic review on traffic-related air pollution. Environ Health Perspect 131:115002, *https://doi.org/10.1289/ehp11532*.

Boogaard H, Crouse DL, Tanner E, Mantus E, van Erp AM, Vedal S, et al. 2024. Assessing adverse health effects of long-term exposure to low levels of ambient air pollution: The HEI experience and what's next? Environ Sci Technol 58:12767–12783, *https://doi.org/10.1021/acs.est.3c09745*.

Brunekreef B, Strak M, Chen J, Andersen ZJ, Atkinson R, Bauwelinck M, et al. 2021. Mortality and Morbidity Effects of Long-Term Exposure to Low-Level $PM_{2.5}$, BC, NO_2, and O_3: An Analysis of European Cohorts in the ELAPSE project. Research Report 208. Boston, MA: Health Effects Institute.

Chen J, Braun D, Christidis T, Cork M, Rodopoulou S, Samoli E, et al. 2023. Long-term exposure to low-level $PM_{2.5}$ and mortality: investigation of heterogeneity by harmonizing analyses in large cohort studies in Canada, United States, and Europe. Environ Health Perspect 131:127003, *https://doi.org/10.1289/ehp12141*.

Chen J, Hoek G. 2020. Long-term exposure to PM and all-cause and cause-specific mortality: A systematic review and meta-analysis. Environ Int 143:105974, *https://doi.org/10.1016/j.envint.2020.105974*.

de Hoogh K, Chen J, Gulliver J, Hoffmann B, Hertel O, Ketzel M, et al. 2018. Spatial $PM_{2.5}$, NO_2, O_3 and BC models for Western Europe: evaluation of spatiotemporal stability. Environ Int 120:81–92, *https://doi.org/10.1016/j.envint.2018.07.036*.

Dockery DW, Pope CA 3rd, Xu X, Spengler JD, Ware JH, Fay ME, et al. 1993. An association between air pollution and mortality in six US cities. N Engl J Med 329:1753–1759, *https://doi.org/10.1056/nejm199312093292401*.

EC (European Commission). 2022. Proposal for a Directive of the European Parliament and the Council on ambient air quality and cleaner air for Europe. Brussels, 26.10.2022. COM (2022) 542 final.

Eeftens M, Beelen R, de Hoogh K, Bellander T, Cesaroni G, Cirach M, et al. 2012. Development of land use regression models for $PM_{2.5}$, $PM_{2.5}$ absorbance, PM_{10} and PM_{coarse} in 20 European study areas: results of the ESCAPE project. Environ Sci Technol 46:11195–11205, *https://doi.org/10.1021/es301948k*.

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

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

HEI Review Panel on Ultrafine Particles. 2013. Understanding the Health Effects of Ambient Ultrafine Particles. Perspectives 3. Boston MA: Health Effects Institute.

Hoek G. 2017. Methods for assessing long-term exposures to outdoor air pollutants. Curr Environ Health Rep 4:450–462, *https://doi.org/10.1007/s40572-017-0169-5*.

Hoek G, Vienneau D, de Hoogh K. 2024. Does residential address-based exposure assessment for outdoor air pollution lead to bias in epidemiological studies? Environ Health 23:75, *https://doi.org/10.1186/s12940-024-01111-0*.

Huangfu P, Atkinson R. 2020. Long-term exposure to NO_2 and O_3 and all-cause and respiratory mortality: a systematic review and meta-analysis. Environ Int 144:105998, *https://doi.org/10.1016/j.envint.2020.105998*.

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

Jerrett M, Arain A, Kanaroglou P, Beckerman B, Potoglou D, Sahsuvaroglu T, et al. 2005. A review and evaluation of intraurban air pollution exposure models. J Exposure Anal Environ Epidemiol 15:185–204, *https://doi.org/10.1038/sj.jea.7500388*.

Kasdagli MI, Orellano P, Pérez Velasco R, Samoli E. 2024. Long-term exposure to nitrogen dioxide and ozone and mortality: update of the WHO air quality guidelines systematic review and meta-analysis. Int J Public Health 69:1607676, *https://doi.org/10.3389/ijph.2024.1607676*.

Kioumourtzoglou MA, Spiegelman D, Szpiro AA, Sheppard L, Kaufman JD, Yanosky JD, et al. 2014. Exposure measurement error in $PM_{2.5}$ health effects studies: a pooled analysis of eight personal exposure validation studies. Environ Health 13:2, *https://doi.org/10.1186/1476-069x-13-2*.

Kerckhoffs J, Hoek G, Messier KP, Brunekreef B, Meliefste K, Klompmaker JO, Vermeulen R. 2016. Comparison of ultrafine particle and black carbon concentration predictions from a mobile and short-term stationary land-use regression model. Environ Sci Technol 50:12894–12902, *https://doi.org/10.1021/acs.est.6b03476*.

Kerckhoffs J, Hoek G, Vlaanderen J, van Nunen E, Messier K, Brunekreef B, et al. 2017. Robustness of intra urban land-use regression models for ultrafine particles and black carbon based on mobile monitoring. Environ Res 159:500–508, *https://doi.org/10.1016/j.envres.2017.08.040*.

Kerckhoffs J, Hoek G, Gehring U, Vermeulen R. 2021. Modelling nationwide spatial variation of ultrafine particles based on mobile monitoring. Environ Int 154:106569, *https://doi.org/10.1016/j.envint.2021.106569*.

Kerckhoffs J, Khan J, Hoek G, Yuan Z, Hertel O, Ketzel M, et al. 2022. Hyperlocal variation of nitrogen dioxide, black carbon, and ultrafine particles measured with Google Street View cars in Amsterdam and Copenhagen. Environ Int 170:107575, *https://doi.org/10.1016/j.envint.2022.107575*.

Klompmaker JO, Janssen N, Andersen ZJ, Atkinson R, Bauwelinck M, Chen J, et al. 2021. Comparison of associations between mortality and air pollution exposure estimated with a hybrid, a land-use regression and a dispersion model. Environ Int 146:106306, *https://doi.org/10.1016/j.envint.2020.106306*.

Orellano P, Kasdagli MI, Pérez Velasco R, Samoli E. 2024. Long-term exposure to particulate matter and mortality: an update of the WHO global air quality guidelines systematic review and meta-analysis. Int J Public Health 69:1607683, *https://doi.org/10.3389/ijph.2024.1607683*.

Pope CA 3rd, Burnett RT, Thun MJ, Calle EE, Krewski D, Ito K, et al. 2002. Lung cancer, cardiopulmonary mortality, and long-term exposure to fine particulate air pollution. JAMA 287:1132–1141, *https://doi.org/10.1001/jama.287.9.1132*.

Samoli E, Butland BK. 2017. Incorporating measurement error from modeled air pollution exposures into epidemiological analyses. Curr Environ Health Rep 4:472–480, *https://doi.org/10.1007/s40572-017-0160-1*.

Sheppard L, Burnett RT, Szpiro AA, Kim SY, Jerrett M, Pope CA 3rd, et al. 2012. Confounding and exposure measurement error in air pollution epidemiology. Air Qual Atmos Health 5:203–216, *https://doi.org/10.1007/s11869-011-0140-9*.

Shen Y, de Hoogh K, Schmitz O, Clinton N, Tuxen-Bettman K, Brandt J, et al. 2022. Europe-wide air pollution modeling from 2000 to 2019 using geographically weighted regression. Environ Int 168:107485, *https://doi.org/10.1016/j.envint.2022.107485*.

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

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

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

US EPA (United States Environmental Protection Agency). 2024. Final rule: Reconsideration of the national ambient air quality standards for particulate matter. Available: *https://www.epa.gov/pm-pollution/final-reconsideration-national-ambient-air-quality-standards-particulate-matter-pm* [accessed November 13, 2024].

Velders GJM, Willers SM, Wesseling J, den Elshout SV, van der Swaluw E, Mooibroek D, et al. 2021. Improvements in air quality in the Netherlands during the corona lockdown based on observations and model simulations. Atmos Environ 247:118158, *https://doi.org/10.1016/j.atmosenv.2020.118158*.

Wesseling J, de Ruiter H, Blokhuis C, Drukker D, Weijers E, Volten H, et al. 2019. Development and implementation of a platform for public information on air quality, sensor measurements, and citizen science. Atmosphere 10:445, *http://dx.doi.org/10.3390/atmos10080445*.

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

Yu Z, Koppelman GH, Boer JMA, Hoek G, Kerckhoffs J, Vonk JM, et al. 2022. Ambient ultrafine particles and asthma onset until age 20: the PIAMA birth cohort. Environ Res 214:113770, *https://doi.org/10.1016/j.envres.2022.113770*.

Yu Z, Koppelman GH, Hoek G, Kerckhoffs J, Vonk JM, Vermeulen R, et al. 2021. Ultrafine particles, particle components and lung function at age 16 years: the PIAMA birth cohort study. Environ Int 157:106792, *https://doi.org/10.1016/j.envint.2021.106792*.