Environmental Research 284 (2025) 122237

Contents lists available at ScienceDirect

Environmental Research

journal homepage: www.elsevier.com/locate/envres

Measurement error correction methods for the effects of ambient air pollution on mortality and morbidity using the UK Biobank cohort: the MELONS study

Dimitris Evangelopoulos [a,b,*,1], Dylan Wood [a,b,1], Barbara K. Butland [c], Benjamin Barratt [a,b], Hanbin Zhang [a,d], Konstantina Dimakopoulou [e], Evangelia Samoli [e], Sean Beevers [a,b], Heather Walton [a,b], Joel Schwartz [f,g], Evangelos Evangelou [h], Klea Katsouyanni [a,b,e]

[a] Environmental Research Group, MRC Centre for Environment and Health, Faculty of Medicine, Imperial College London, United Kingdom
[b] NIHR HPRU in Environmental Exposures and Health, Imperial College London, United Kingdom
[c] School of Health & Medical Sciences, City St George's, University of London, United Kingdom
[d] European Centre for Environment and Human Health, University of Exeter, United Kingdom
[e] Department of Hygiene, Epidemiology and Medical Statistics, National and Kapodistrian University of Athens, Medical School, Greece
[f] Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, MA, USA
[g] Department of Environmental Health, Harvard T.H. Chan School of Public Health, Boston, MA, USA
[h] Department of Epidemiology and Biostatistics, School of Public Health, Imperial College London, United Kingdom

ARTICLE INFO

Keywords:
Air pollution
Exposure measurement error
Long-term exposure
Personal exposure
Mortality
Morbidity

ABSTRACT

Epidemiological cohort studies associating long-term exposure to ambient air pollution with health outcomes most often do not account for individually assigned exposure measurement error. Here, we implemented Cox proportional hazards models to explore the relationships between NO_2, $PM_{2.5}$ and ozone exposures with the incidence of natural-cause mortality and several morbidity outcomes in 61,797 London-dwelling respondents of the UK Biobank cohort. Data from an existing personal monitoring campaign was used as an external validation dataset to estimate measurement error structures between "true" personal exposure and several surrogate (measured and modelled) estimates of assigned exposure, allowing for the application of two health effect estimate correction methodologies: regression calibration (RCAL) and simulation extrapolation (SIMEX). Uncorrected hazard ratios (HRs) suggested an increase in the risk of natural-cause mortality for modelled NO_2 estimates (HR: 1.028 [0.983, 1.074] per IQR increment of 14.54 μg/m^3) and no statistically significant association was observed for $PM_{2.5}$ surrogate exposure measures. Measurement error corrected HRs were generally larger in magnitude, although exhibited wider confidence intervals than uncorrected effect estimates. Chronic obstructive pulmonary disease (COPD) was associated with increased exposure to modelled NO_2 (1.087 [1.022, 1.155]). Both RCAL and SIMEX correction resulted in increased HRs (1.254 [1.061, 1.482] and 1.192 [1.093, 1.301], respectively). SIMEX correction of modelled $PM_{2.5}$ (IQR: 1.72 μg/m^3) associations with COPD increased the HR (1.079 [1.001, 1.164]) in comparison to uncorrected (1.042 [0.988, 1.099]). These findings suggest that health effect estimates not corrected for exposure measurement error may lead to underestimation in the magnitude of effects.

1. Introduction

Numerous epidemiological studies have documented associations between long-term ambient air pollution exposure and mortality (Chen and Hoek, 2020; Huangfu and Atkinson, 2020), as well as morbidity outcomes, including cardiorespiratory illnesses (Forastiere et al., 2024).

* Corresponding author. Environmental Research Group, MRC Centre for Environment and Health, Faculty of Medicine, Imperial College London, White City Campus, Wood Lane, London, W12 0BZ, United Kingdom
E-mail address: d.evangelopoulos@imperial.ac.uk (D. Evangelopoulos).
[1] Joint first authors.

https://doi.org/10.1016/j.envres.2025.122237
Received 19 March 2025; Received in revised form 5 June 2025; Accepted 24 June 2025
Available online 28 June 2025

A crucial issue in such analyses is the assignment of air pollution concentrations as a measure of exposure. Most studies assign long-term average concentrations of ambient air pollutants to individuals using measurements from fixed-site monitoring networks or estimates derived from other approaches, such as dispersion or land-use regression models, machine learning approaches including satellite data or combinations of these methodologies (Steinle et al., 2013; Shen et al., 2024). Concentration estimates derived using such methods are usually assigned to individuals in cohort studies at the residential postal/zip code level, aiming to investigate the associations between increased long-term exposure and health endpoints. Generally, concentrations are averaged over a period of time. However, the assignment of such estimates may not accurately reflect actual individual-level personal exposure to ambient air pollution, given the amount of time people spend within indoor environments and travelling through different transport microenvironments (Evangelopoulos et al., 2020). Error in the assigned exposure in epidemiological studies is referred to as exposure measurement error and likely leads to downward bias for resulting health effect estimates (Samoli et al., 2020; Butland et al., 2020; Wei et al., 2022).

In an epidemiological analysis, the type and magnitude of measurement error in the exposure can result in varying biases (Katsouyanni and Evangelopoulos, 2022). Classical error assumes that the surrogate exposure is an imperfect estimate of the true exposure, and the average of these surrogate estimates equals the true exposure. Measurement error is uncorrelated with the true exposure under the classical error model. Berkson error occurs when the average value of the true exposure equals the surrogate estimate and applies when the surrogate represents a shared exposure across a group of participants whose individual exposures might differ. Zeger et al. (2000) suggest that classical error is observed when you compare measured ambient levels of pollution and the true values for a measuring device that is unbiased (Zeger et al., 2000). Berkson error would be observed when you compare ambient levels from a reference monitor against multiple measurements from low-cost sensors located around the reference monitor.

Additive classical measurement error generally biases health effect estimates towards the null and decreases the coverage of 95 % confidence intervals, whereas additive Berkson error tends to inflate standard errors and, thus, reduce statistical power (Carroll et al., 2006), although there are exceptions to these "general" statements (Samoli et al., 2020; Butland et al., 2020). In environmental health studies, modelled ambient air pollution concentrations are prone to complex random measurement error involving mixtures of both classical/classical-like error from parameter estimation in the exposure approaches and Berkson/Berkson-like error from over-smoothing (Zeger et al., 2000). As a result of these errors in exposure estimation, it may be difficult to interpret health effect estimates reported for a given increment in ambient pollutant concentrations, particularly when comparing studies utilising various exposure assignment methodologies. Such errors likely lead to heterogeneity in effect estimates (e.g. (Strak et al., 2021; Wolf et al., 2023),). Furthermore, the likely presence and unknown magnitude of exposure measurement error in the air pollution epidemiology literature presents further challenges when collating the existing evidence to provide accurate concentration-response functions (CRFs) in health impact assessments relevant for actionable policy (Forastiere et al., 2024). A meta-regression to estimate a CRF found larger effect sizes in studies with less exposure error but did not estimate the effect of the error on the shape of the CRF (Vodonos et al., 2018).

Estimating accurate "true" personal exposure to ambient air pollution is inherently difficult. Large-scale personal monitoring campaigns are very limited to date, given the financial cost, as well as the inconvenience to participants themselves. For studies of the long-term effects of air pollution, personal monitoring for years is infeasible. A recent systematic review comparing static exposure assessment with exposures that integrate time-activity patterns showed high correlations between exposures, small to moderate bias in health effect estimates and modest differences in exposure level and contrast (Hoek et al., 2024). However, the review was based on a small number of identified studies (only 11) which performed indirect adjustment of exposure using time-activity data, rather than measuring exposures with portable monitors. In a similar analysis in London, we observed a consistent pattern of slightly increased adverse effect estimates of time activity-adjusted air pollution exposure with executive function scores (i.e., a domain-specific measure of cognitive function) compared to static exposures (Wood et al., 2025). It should be noted that personal exposure to ambient air pollution cannot be measured directly but can be estimated from total personal exposure measurements after the separation of indoor- and outdoor-generated pollution (Evangelopoulos et al., 2024). The processes required in conducting such separation entail several assumptions.

The identification of the health impacts of air pollution on mortality and morbidity in the UK has been the aim of multiple previous studies, with many applications based on the UK Biobank cohort dataset (e.g. (Hansell et al., 2024),). These studies used exposures developed by multiple research groups, were based on different approaches and showed adverse effects on lung function, cardiovascular disease and mortality associated with increased exposures to particulate matter up to 2.5 and 10 μm in diameter, as well as nitrogen dioxide ($PM_{2.5}$, PM_{10} and NO_2, respectively (Wang et al., 2022; Guyatt et al., 2024);). However, no study to our knowledge has assessed measurement error bias in these estimates or implemented a measurement error correction method.

The present study utilised data from an existing personal monitoring campaign to estimate the magnitude of measurement error present when assigning measured and modelled proxies of personal exposure to the same individuals. This information on measurement error structures was then applied in two separate correction methods, i.e., regression calibration (RCAL) and simulation extrapolation (SIMEX), aiming to adjust health effect estimates associating air pollution exposure with natural-cause mortality, chronic obstructive pulmonary disease (COPD) incidence and other morbidity outcomes in the UK Biobank cohort. The present study was conducted as part of the US Health Effects Institute-funded project entitled "Investigating the Consequences of Measurement Error of Gradually More Sophisticated Long-Term Personal Exposure Models in Assessing Health Effects: The London Study" (the MELONS project).

2. Methods

2.1. Study population

London-dwelling participants of the UK Biobank cohort (www.ukbiobank.ac.uk) were included in this analysis in order to match our exposure assessment methods with the personal monitoring campaign. UK Biobank is a national cohort of more than 500,000 participants aged 40–69 years when recruited between 2006 and 2010 from general practice registers. The study sample consisted of 62,029 participants residing in the Greater London area, who were recruited between 13 June 2006 and 24 September 2010, and followed up until 08 January 2024. Of this sample, 0.4 % (n = 232) were initially excluded from the analysis as they were lost to follow-up for the following reasons: participants left the UK (n = 202), they withdrew consent for the UK Biobank database (n = 27), or their death was reported by a relative without any further information on the exact date and cause (n = 3). Thus, the final number of participants included in the analysis was 61,797 Greater London residents.

2.2. Health outcome data

Health outcomes of focus were the number of cases for mortality and COPD incidence, but we also examined other morbidity outcomes. Natural-cause mortality (ICD-10 codes: A00-R99) was defined based on

the primary cause of death recorded in mortality registries. Incident cases of COPD (J43-J44) were derived from linked hospital episode statistics data in the UK Biobank cohort. Additionally, incident cases of myocardial infarction (MI, I21-I23, I24.1, I25.2), asthma (J45-J46), stroke (I60-I61, I63-I64) and all-cause dementia (A81.0, F00-F03, F05.1, F10.6, G30-G31) were similarly derived. Prevalent cases at baseline for each outcome were excluded from the relevant analyses. The end of follow-up was defined as the date of death or diagnosis of outcome for the cases, the date we received the final dataset (08 January 2024) for those still alive or the date of death from external causes for censored observations.

2.3. Exposure assessment

2.3.1. Surrogate measures of exposure

We assigned long-term estimates of exposure to NO_2, $PM_{2.5}$ and ozone (O_3) to London-dwelling UK Biobank participants, including predictions from spatio-temporal models, concentrations measured at the nearest fixed-site monitor and hybrid methods accounting for people's mobility. Specifically, annually averaged 1-h resolution background and roadside site measurements of NO_2, $PM_{2.5}$ and O_3 from the London Air Quality Network (LAQN) monitors (within the Greater London area) were assigned to each participant for the year of recruitment into the UK Biobank study. These publicly available measurements were downloaded via the '*openair*' R package (Carslaw and Ropkins, 2012). Measurements from urban background and roadside monitors of the air quality network were used, with kerbside and industrial sites excluded.

Modelled NO_2, $PM_{2.5}$ and O_3 concentrations were derived using the model from the "Comparative evaluation of Spatio-Temporal Exposure Assessment Methods for estimating the health effects of air pollution" (STEAM) project (long-term 2009–2013 averaged estimates (Dimakopoulou et al., 2022);) and assigned to the residential address of UK Biobank participants (within a 100m buffer of the home) at the time of recruitment. The STEAM model utilised a generalised additive model that combined data from spatio-temporal land-use regression and the Community Multiscale Air Quality urban (CMAQ-urban (Beevers et al., 2012);) chemical transport model and, additionally for $PM_{2.5}$ only, a prediction model using satellite-data on aerosol optical depth (Samoli et al., 2020; Butland et al., 2020). Models provided daily predictions, which were averaged annually and then for 2009–2013 to assess long-term exposures. The 10-fold cross validation R^2 were 0.80 for NO_2, 0.79 for $PM_{2.5}$ and 0.75 for O_3 (Dimakopoulou et al., 2022).

Modelled and measured estimates of NO_2 and $PM_{2.5}$ were also indirectly adjusted to account for estimated time-activity patterns of UK Biobank participants by age group and area of residence. Specifically, we utilised the London Hybrid Exposure Model (LHEM (Smith et al., 2016)), which combines time-activity patterns of Londoners with CMAQ-urban modelled ambient NO_2 and $PM_{2.5}$ concentrations (in 2011) with in-building and in-vehicle modelled concentrations. The LHEM estimates personal exposure to outdoor-generated pollution reflective of time-activity and microenvironment exposure (Smith et al., 2016). The model incorporated data from the London Travel Demand Survey (LTDS), which includes 69,673 individuals across the years 2005–2010, ascertaining details of each persons' daily trips, travel mode (s), trip purposes and demographic data. These individuals are a representative sample of London inhabitants. The ratio between the static concentrations estimated by CMAQ-urban at the residential address of LTDS participants and their LHEM personal exposure, accounting for time-activity and movement between microenvironments, were applied to modelled and measured exposure estimates assigned to UK Biobank participants to account for their hypothesised time-activity patterns. Further details on the time-activity adjustment can be found in (Wood et al., 2025).

2.3.2. "True" exposure derived from a personal monitoring campaign

Data from a previous exposure measurement campaign conducted in London between 2015 and 2017 for 71 COPD patients aged 53 years and older were utilised as an external validation sample. Total personal exposure measurements for NO_2, $PM_{2.5}$ and O_3, as well as GPS data measured by portable monitors for up to six months, were used. Further details on the "Characterisation of COPD exacerbations using environmental exposure modelling" study (COPE) can be found in (Moore et al., 2016).

Long-term estimates of exposure to ambient air pollution were assumed as "true" personal exposure to each pollutant for COPE participants. The derivation of these estimates is described in (Barratt et al., 2022). Briefly, previous work (as part of the MELONS project) classified the location of the participants into microenvironments including 'home', 'other indoor', and 'outdoor/transit', using 15-min resolution GPS data. Personal air pollution exposures were also collected at 15-min intervals and averaged at 1-h resolution, while the location assigned was the one within which the participant spent the most time across the 1-h period. Total measured personal exposure to each pollutant was separated into exposure from indoor and outdoor sources (Evangelopoulos et al., 2024; Zhang et al., 2022) and outdoor-generated personal exposures were extrapolated to annual average using a random forest behaviour prediction model.

2.4. Statistical analysis

Cox proportional hazards models were applied to quantify the associations between the incidence of each health outcome and exposure to NO_2, $PM_{2.5}$ and O_3 derived from the aforementioned surrogate exposure metrics in single pollutant models without measurement error correction. Three complete case Cox models adjusting for confounders were constructed *a priori*, with UK Biobank participants missing any confounder information excluded. A basic model (Model 1) was adjusted for age (in years), sex (Male/Female) and year of recruitment (calendar year), with length of time from date of recruitment up until the date of outcome diagnosis, loss to follow-up or the end of the study period (08 January 2024) used as the underlying measure of time. Model 2, *a priori* considered as our "main model", further controlled for ethnic background (Asian, Black/Black British, Mixed, White or 'other'), smoking status (Current/Previous/Never), BMI (in kg/m^2), household income (categorised in GB pounds (£) bands, as $<$18,000, [18,000–31,000), [31,000–52,000), [52,000–100,000), $\geq$100,000) and employment status (Employed/Unemployed/Retired/Other; including "doing unpaid or voluntary work", "full- or part-time student", "looking after home and/or family", "unable to work because of sickness or disability"). Model 3 further included the index of multiple deprivation (IMD) score as a marker of area-level socioeconomic status. IMD score is comprised of eight factors of deprivation at neighbourhood level, including living environment (Noble et al., 2019). To avoid double adjustment for air pollution exposure, Model 3 was included as part of our sensitivity analyses. All hazard ratios were estimated per interquartile range (IQR) increase in pollutant concentrations (in $\mu g/m^3$) and the threshold used for determining statistical significance was 0.05.

We applied two measurement error corrections on Models 1 and 2 using RCAL and SIMEX. In brief, RCAL uses the linear relationship between "true" and surrogate exposure estimates in a validation subsample in which accurate, "gold-standard" measurements are available (the RCAL model) to predict "true" exposure data in the main study of cohort participants. The predicted "true" exposure estimates are then used in the full cohort analysis to estimate measurement error bias-corrected health effect estimates, assuming the "true"-surrogate exposure relationship in the validation sub-sample generalises to the full cohort. SIMEX uses simulation to explore and model changes in the log hazard ratio (HR) of interest as we increase the amount of classical error in the surrogate exposure data. A SIMEX tuning parameter "lambda" is used iteratively to adjust the levels of measurement error in the exposure

before back-extrapolating to the "no error" case and a curve is fitted to describe the back-extrapolated relationship (the extrapolant function). The "no error" projected value provides an adjusted estimate that corrects for bias introduced by the original amount of classical measurement error. Further details on both measurement error correction methods can be found in (Carroll et al., 2006).

Model 2 results were corrected as the main model with more individual-level confounder control. Model 1 HRs were also corrected using RCAL because confounder adjustment in Model 1 corresponded to the variables included in the RCAL model. No internal subsample validation dataset was available in the UK Biobank, so we used the COPE measurement campaign dataset (personal "true" exposure described above) to inform the measurement error structures for correction. Both COPE and UK Biobank participants included in the present study resided in the Greater London area but did not overlap. Natural-cause mortality and COPD incidence were used as the primary outcomes of interest.

The present study assumed that the exposure errors observed were of additive mixed (classical and Berkson) type. For RCAL, calibration coefficients were first estimated from a linear model of the assumed "true" exposure (annually extrapolated personal exposure from outdoor sources derived from COPE) on each surrogate (modelled and measured concentration estimates as described above, which were also assigned to COPE individuals at the residential postcode centroid), controlling for age and sex. Regression calibration was implemented by using the RCAL model to predict corrected exposures for the study subjects and these corrected exposures were then used in the Cox analysis to obtain revised HRs. The standard errors derived from the Cox models were then used to estimate 95 % confidence intervals. As a sensitivity analysis, we calculated standard errors for the health effect estimates using the Delta method and incorporated the uncertainty from the calibration coefficients into the exposure-outcome model (Carroll et al., 2006). The Delta method accounts for the standard error of the pollutant parameter in the RCAL model when estimating the standard error of the RCAL-adjusted HR (Hart et al., 2015a). For SIMEX, we followed the methods described in (Reeves et al., 1998). Briefly, the additive classical error variance from the COPE validation dataset was estimated assuming the latent variable mixed error model, which represents "true" exposure by an unobserved latent variable plus Berkson measurement error, and the corresponding surrogate exposure by the same latent variable plus classical measurement error. The latent variable mixed error model provides a flexible framework through which we estimated the proportion of classical and Berkson error in the total mixture for each pollutant independently. This was expressed by the variance ratio of surrogate over "true" exposure, with ratios >1 indicating mostly classical error and ratios <1 mostly Berkson-like error. The model assumes that the three variables (latent, classical and Berkson) are assumed to be independent or at least mutually uncorrelated (Reeves et al., 1998). The SIMEX lambda parameter was set to take 20 values with a maximum value of 2 and 50 simulations performed for each lambda. We also assessed the sensitivity of results to these inputs. Jacknife estimated standard errors were calculated and the extrapolant used was the quadratic. SIMEX was implemented using the "*simex*" package in R (Lederer et al., 2019).

3. Results

Table 1 provides summary statistics for surrogate residential air pollution exposure estimates assigned to UK Biobank participants. STEAM model and nearest monitor mean (standard deviation) levels were relatively high for all three pollutants included in the analysis, i.e., 39.0 (10.8) and 46.8 (14.8) μg/m^3 for NO_2, 15.8 (1.4) and 15.2 (1.9) μg/m^3 for $PM_{2.5}$, as well as 58.8 (6.8) and 34.0 (6.5) μg/m^3 for O_3, respectively. The LHEM time-activity adjustment resulted in substantially lower mean exposures for both NO_2 and $PM_{2.5}$. Moderate correlation was observed between NO_2 and $PM_{2.5}$, dependent upon the surrogate measure, ranging from 0.30 (STEAM) to 0.44 (nearest monitor). Ozone was negatively correlated with both NO_2 (−0.99 and −0.63 for STEAM and nearest monitor estimates, respectively), as well as with $PM_{2.5}$ (−0.29 and −0.27 for STEAM and nearest monitor estimates, respectively).

Table 2 provides data for potential confounders, number of deaths from natural causes (n = 4,138), total person-years at risk (847,455.5) and the number of COPD incident cases (n = 2,081) observed throughout follow-up in the Greater London-dwelling UK Biobank participants. The mean (SD) age of participants was 55.9 (8.3) years, mean BMI was 27.0 (4.9) kg/m^2, 55.8 % were female, 79.7 % were white and 44.6 % were current or previous smokers. Incident cases for the other health outcomes were less than COPD cases and ranged from 1,044 for all-cause dementia to 1,677 for MI. Missing individual-level covariate information was relatively low and resulted in approximately 2 % of individuals being removed from main model analyses of natural-cause mortality and COPD (Model 2).

For the external validation sample of COPE participants, the surrogate measures assessing ambient concentrations at the subjects' residence (both modelled and measured at the nearest monitor) were much higher in magnitude compared to personal exposures from outdoor sources (Table 3). The LHEM time-activity-adjusted estimates were closer to measured personal exposures to air pollution from outdoor sources. Compared with the exposure assessment of the UK Biobank participants, all surrogate exposure estimates of the validation sample were very similar, although greater variability was observed in STEAM NO_2 estimates assigned to COPE participants. For example, the mean (standard deviation) STEAM estimates for COPE were 40.1 μg/m^3 (18.7), 15.6 μg/m^3 (1.4) and 60.0 μg/m^3 (7.2) for NO_2, $PM_{2.5}$ and O_3,

Table 1
Descriptive statistics for surrogate exposure variables for the 61,797 participants of the UK Biobank cohort residing in the Greater London area.

Exposure (μg/m^3)	Mean	SD	Minimum	25th %ile	Median	75th %ile	Maximum	IQR
NO_2								
STEAM	38.99	10.79	17.60	31.05	38.16	45.59	90.50	14.54
STEAM LHEM-adjusted	14.63	3.76	5.96	12.10	14.34	16.81	44.61	4.71
Nearest monitor (annual)	46.84	14.84	21.15	33.37	49.77	55.79	107.32	22.42
Nearest monitor LHEM-adjusted	17.63	5.44	6.94	13.48	18.23	20.94	54.20	7.46
$PM_{2.5}$								
STEAM	15.77	1.37	10.89	14.80	15.62	16.52	25.71	1.72
STEAM LHEM-adjusted	9.86	1.04	6.81	9.17	9.74	10.37	25.67	1.20
Nearest monitor (annual)	15.22	1.90	10.95	14.00	15.03	15.90	25.23	1.90
Nearest monitor LHEM-adjusted	9.51	1.22	6.51	8.72	9.42	9.97	21.55	1.25
Ozone								
STEAM	58.79	6.81	34.70	54.26	59.34	63.96	71.46	9.70
Nearest monitor (annual)	33.97	6.50	18.66	28.99	34.24	38.94	51.56	9.95

Table 2
Descriptive statistics for individual- and area-level demographic and socioeconomic variables and health outcomes, for the 61,797 participants of the UK Biobank cohort residing in the Greater London area.

Variable		Number of missing values
Age at recruitment in years (Mean (SD))	55.9 (8.3)	0
Sex (n, %)		0
Male	27,332 (44.2)	
Female	34,465 (55.8)	
BMI in kg/m^2 (Mean (SD))	27.0 (4.9)	701
Area-level IMD score (Mean (SD))	20.1 (12.1)	1,687
Smoking status (n, %)		349
Current	7,348 (12.0)	
Previous	21,258 (34.6)	
Never	32,504 (52.9)	
Prefer not to answer	302 (0.5)	
Employment status (n, %)		295
Employed	35,504 (57.7)	
Unemployed	2,066 (3.4)	
Retired	18,019 (29.3)	
Other	5,916 (9.6)	
Ethnicity (n, %)		351
Asian	4,384 (7.1)	
Black/Black British	4,113 (6.7)	
Mixed	939 (1.5)	
Other	3,037 (5.0)	
White	48,973 (79.7)	
Average household income (n, %)		649
Less than 18,000£	9,736 (15.9)	
18,000 to 30,999£	10,863 (17.8)	
31,000 to 51,999£	12,698 (20.8)	
52,000 to 100,000£	12,741 (20.8)	
Greater than 100,000£	5,938 (9.7)	
Do not know/Prefer not to answer	9,172 (15.0)	
Natural-cause mortality (n, %)	4,138 (6.7)	
COPD incidence (n, %)	2,081 (3.4)	
Myocardial infarction incidence (n, %)	1,677 (2.7)	
Stroke incidence (n, %)	1,266 (2.0)	
All-cause dementia incidence (n, %)	1,044 (1.7)	
Asthma incidence (n, %)	1,597 (2.6)	
Person years at risk (Total)	847,455.5	

respectively, while the corresponding UK Biobank estimates were 39.0 μg/m^3 (10.8), 15.8 μg/m^3 (1.4) and 58.8 μg/m^3 (6.8). Also, the mean multiple deprivation score (IMD) was very similar between the two samples (COPE: 20.4 (11.3), UK Biobank: 20.1 (12.1)), as well as the female/male ratio (COPE: 50.7 % female (n = 38), UK Biobank: 55.8 % female (n = 34,465)). COPE participants were older on average than the UK Biobank participants included in our analysis, i.e., 70.5 (7.5) and 55.9 (8.3) years of age, respectively. The variance ratios between surrogate exposures and outdoor-generated personal exposure ("true") were >1 for NO_2 and O_3, and <1 for $PM_{2.5}$ (except nearest monitor), indicating more classical-like error for the former and Berkson-like for the latter. The Pearson correlation coefficients between surrogate and true exposures were generally low (Table 3). Regression coefficients for each pollutant's surrogate on true exposure from the COPE RCAL models are provided in Supplementary Table 8.

Results from the single exposure Cox regression models for natural-cause mortality are shown in Table 4. RCAL-corrected hazard ratios for Models 1 and Model 2 are shown, as well as SIMEX-corrected HRs for Model 2. The uncorrected HRs suggest an increase in the risk of mortality from natural causes for the STEAM (main and LHEM-adjusted) NO_2 estimates with Models 1 and 2 (for Model 2 it did not reach the nominal level of statistical significance). The measurement error corrected HRs were larger in magnitude, and for the STEAM estimates were more than double in size. For SIMEX specifically, the corrected HR for LHEM-adjusted STEAM was statistically significant, i.e., 1.066 (1.004, 1.131), compared to 1.032 (0.991, 1.075) per 4.71 μg/m^3 increase, under the fully adjusted Model 2. Confidence intervals generally widened after correction in comparison to uncorrected effect estimates, suggesting a higher level of uncertainty in associations even in instances where correction greatly increased the magnitude of the effect estimate and/or resulted in statistical significance. Additionally, calculation of standard errors for RCAL-corrected HRs using the Delta method resulted in comparable, and in some cases widened, CIs (Supplementary Material Table S7).

For $PM_{2.5}$, no association was statistically significant at the nominal level, however the LHEM-adjusted STEAM estimates showed an increased risk for mortality which was not statistically significant. In contrast, the nearest monitor surrogate provided associations close to null and were not statistically significant. The RCAL- and SIMEX-corrected estimates were again larger in size compared to the uncorrected estimates, but not statistically significant. The SIMEX-corrected HR for LHEM-adjusted STEAM was not statistically significant (1.042 (0.989, 1.098) per 4.71 μg/m^3 increase).

For O_3, reverse associations were observed, i.e., protective effects, but only the basic model with limited adjustment (Model 1) using the STEAM exposure estimates was statistically significant. Interestingly, this association became positive, i.e., suggesting adverse effects, after RCAL correction, and remained positive but not statistically significant in the fully adjusted RCAL-corrected Model 2. As expected, the RCAL confidence intervals were generally wider than those for the uncorrected and SIMEX-corrected HRs, due to the type of standard error estimation. To test the assumption of proportionality we assessed the Schoenfeld residuals for each covariate visually. No indication of heteroscedasticity was observed.

Similar results to the measurement error corrected natural-cause mortality estimates were observed for COPD incidence (Table 5). Both NO_2 and $PM_{2.5}$ were associated with an increased risk in COPD incidence based on all uncorrected Models 1 and 2 (except for LHEM-adjusted

Table 3
Descriptive statistics of outdoor-generated personal exposure ("true") and assigned surrogate concentration estimates for the participants of the external validation sample (COPE; n = 71). **VR**: Variance ratio between surrogate and "true" exposure. **Corr**: Pearson correlation coefficient between surrogate and "true" exposures.

Exposure estimate	NO_2			$PM_{2.5}$			O_3		
	Mean (SD)	VR	Corr	Mean (SD)	VR	Corr	Mean (SD)	VR	Corr
Outdoor-generated personal exposure	4.5 (1.5)	–	–	5.6 (2.0)	–	–	2.9 (1.1)	–	–
STEAM	40.1 (18.7)	155.4	0.16	15.6 (1.4)	0.49	0.04	60.0 (7.2)	42.8	−0.21
STEAM LHEM-adjusted	13.5 (4.3)	8.2	0.25	9.4 (1.0)	0.25	0.08	LHEM not applicable		
Nearest monitor (annual)	44.5 (18.1)	145.6	0.13	12.6 (2.3)	1.32	0.10	32.6 (7.9)	51.6	−0.01
Nearest monitor LHEM-adjusted	15.7 (6.5)	18.8	0.08	7.6 (1.4)	0.49	0.09	LHEM not applicable		

Table 4
Investigation of the association between natural-cause mortality and air pollution exposures (entered alternatively), adjusting for confounders using three models with different levels of adjustment. Results from Cox proportional hazards regression models. ME corrected HRs are also presented. Results presented per IQR increase in exposure[c].

Exposure ($\mu g/m^3$)	Hazard ratio per IQR (95 % confidence interval)				
	Model 1[a] (n = 61,797)	RCAL-corrected HR for Model 1	Model 2[b] (n = 60,528)	RCAL-corrected HR for Model 2	SIMEX-corrected HR for Model 2
NO_2					
STEAM	1.061 (1.017, 1.107)*	1.144 (1.017, 1.288)*	1.028 (0.983, 1.074)	1.077 (0.955, 1.215)	1.063 (0.998, 1.134)
STEAM LHEM-adjusted	1.045 (1.006, 1.087)*	1.068 (0.998, 1.143)	1.032 (0.991, 1.075)	1.055 (0.985, 1.131)	1.066 (1.004, 1.131)*
Nearest monitor (annual)	1.012 (0.966, 1.059)	1.021 (0.930, 1.121)	0.992 (0.945, 1.040)	0.983 (0.894, 1.081)	0.990 (0.923, 1.063)
Nearest monitor LHEM-adjusted	0.998 (0.957, 1.041)	1.000 (0.865, 1.155)	0.994 (0.952, 1.038)	0.979 (0.846, 1.134)	0.994 (0.932, 1.060)
$PM_{2.5}$					
STEAM	1.021 (0.983, 1.060)	1.039 (0.908, 1.188)	0.998 (0.959, 1.037)	0.991 (0.865, 1.136)	1.001 (0.946, 1.059)
STEAM LHEM-adjusted	1.021 (0.985, 1.059)	1.144 (0.877, 1.494)	1.022 (0.985, 1.060)	1.166 (0.894, 1.522)	1.042 (0.989, 1.098)
Nearest monitor (annual)	0.989 (0.956, 1.022)	0.898 (0.731, 1.105)	0.983 (0.950, 1.017)	0.901 (0.731, 1.111)	0.973 (0.927, 1.022)
Nearest monitor LHEM-adjusted	0.992 (0.959, 1.025)	0.928 (0.717, 1.201)	1.002 (0.968, 1.037)	1.014 (0.782, 1.315)	1.008 (0.959, 1.060)
Ozone					
STEAM	0.939 (0.898, 0.983)*	1.071 (1.009, 1.137)*	0.969 (0.924, 1.016)	1.042 (0.980, 1.107)	0.949 (0.885, 1.017)
Nearest monitor (annual)	0.998 (0.953, 1.046)	0.989 (0.423, 2.312)	1.009 (0.962, 1.059)	0.849 (0.356, 2.022)	1.024 (0.954, 1.100)

*p < 0.05.
NO_2: STEAM = 14.54, STEAM LHEM-adjusted = 4.71, Nearest monitor (annual) = 22.42, Nearest monitor LHEM-adjusted = 7.46.
$PM_{2.5}$: STEAM = 1.72, STEAM LHEM-adjusted = 1.20, Nearest monitor (annual) = 1.90, Nearest monitor LHEM-adjusted = 1.25.
Ozone: STEAM = 9.70, Nearest monitor (annual) = 9.95.
[a] adjusting for age, time from date of recruitment (timescale variable) and sex.
[b] Model 1 adjustments, plus smoking status, BMI, employment status, ethnicity, household income.
[c] IQRs.

Table 5
Investigation of the association between COPD incidence and air pollution exposures (entered alternatively), adjusting for confounders using three models with different levels of adjustment. Results from Cox proportional hazards regression models. ME corrected HRs are also presented. Results presented per IQR increase in exposure[c].

Exposure ($\mu g/m^3$)	Hazard ratio per IQR (95 % confidence interval)				
	Model 1[a] (n = 61,797)	RCAL-corrected HR for Model 1	Model 2[b] (n = 60,528)	RCAL-corrected HR for Model 2	SIMEX-corrected HR for Model 2
NO_2					
STEAM	1.164 (1.098, 1.234)*	1.506 (1.280, 1.773)*	1.087 (1.022, 1.155)*	1.254 (1.061, 1.482)*	1.192 (1.093, 1.301)*
STEAM LHEM-adjusted	1.096 (1.039, 1.157)*	1.166 (1.062, 1.281)*	1.051 (0.993, 1.112)	1.088 (0.988, 1.198)	1.101 (1.016, 1.194)*
Nearest monitor (annual)	1.073 (1.007, 1.145)*	1.142 (1.001, 1.302)*	1.017 (0.951, 1.088)	1.035 (0.904, 1.184)	1.039 (0.943, 1.146)
Nearest monitor LHEM-adjusted	1.022 (0.963, 1.084)	1.061 (0.865, 1.300)	0.992 (0.933, 1.055)	0.973 (0.791, 1.198)	0.991 (0.906, 1.083)
$PM_{2.5}$					
STEAM	1.075 (1.021, 1.132)*	1.304 (1.084, 1.569)*	1.042 (0.988, 1.099)	1.155 (0.958, 1.394)	1.079 (1.001, 1.164)*
STEAM LHEM-adjusted	1.031 (0.980, 1.084)	1.285 (0.887, 1.863)	1.028 (0.977, 1.081)	1.222 (0.847, 1.762)	1.052 (0.981, 1.129)
Nearest monitor (annual)	1.016 (0.971, 1.063)	1.169 (0.877, 1.559)	1.010 (0.964, 1.058)	1.065 (0.796, 1.424)	1.023 (0.958, 1.094)
Nearest monitor LHEM-adjusted	0.989 (0.944, 1.036)	0.991 (0.69, 1.422)	1.003 (0.956, 1.052)	1.021 (0.711, 1.465)	1.009 (0.943, 1.080)
Ozone					
STEAM	0.850 (0.798, 0.905)*	1.232 (1.133, 1.339)*	0.913 (0.855, 0.975)*	1.125 (1.033, 1.225)*	0.851 (0.774, 0.936)*
Nearest monitor (annual)	0.967 (0.906, 1.032)	1.829 (0.550, 6.081)	1.002 (0.935, 1.074)	0.964 (0.278, 3.339)	1.010 (0.914, 1.117)

*p < 0.05.
NO_2: STEAM = 14.54, STEAM LHEM-adjusted = 4.71, Nearest monitor (annual) = 22.42, Nearest monitor LHEM-adjusted = 7.46.
$PM_{2.5}$: STEAM = 1.72, STEAM LHEM-adjusted = 1.20, Nearest monitor (annual) = 1.90, Nearest monitor LHEM-adjusted = 1.25.
Ozone: STEAM = 9.70, Nearest monitor (annual) = 9.95.
[a] adjusting for age, time from date of recruitment (timescale variable) and sex.
[b] Model 1 adjustments, plus smoking status, BMI, employment status, ethnicity, household income.
[c] IQRs.

nearest monitor estimates of $PM_{2.5}$), with the opposite observed for O_3. The association was statistically significant for STEAM estimated NO_2 after full adjustment, with a HR of 1.087 (1.022, 1.155) per 14.54 $\mu g/m^3$ increase. Again, widened CI ranges were observed after effect estimate correction.

In all statistically significant uncorrected effect estimates, the HRs remained statistically significant after SIMEX or RCAL correction, the magnitude of the effect of an IQR increase in pollutant concentrations increased and CIs widened. The largest statistically significant effects were observed for NO_2, for which the SIMEX- and RCAL-corrected HRs suggested a 19.2 % (9.3, 30.1) and 25.4 % (6.1, 48.2) increase, respectively, in the risk of developing COPD per IQR increase in concentrations (more than a two- or three-fold increase in that observed in the uncorrected main adjustment model; 8.7 % increase in risk). Finally, the O_3 RCAL-corrected estimates showed statistically significant adverse effects, i.e., HR of 1.125 (1.033, 1.225) per 9.70 $\mu g/m^3$ increase for STEAM estimates in the fully adjusted model, in contrast with the uncorrected (0.913 (0.855, 0.975)) and SIMEX-corrected (0.851 (0.774, 0.936)) HRs.

While purely Berkson error would not result in bias in the effect estimates and would only inflate the confidence intervals, even small percentages of classical error in the mixture can introduce bias in the health effect estimates. This was observed for $PM_{2.5}$ which had variance ratios <1, and as low as 0.25 for LHEM-adjusted STEAM. The SIMEX corrected estimates were higher than the naive ones, but the 95 % confidence intervals were inflated although not substantially.

The results observed for the other health outcomes (incidence of myocardial infarction, asthma, stroke and all-cause dementia) generally showed similar patterns in the corrected health effect estimates across surrogate exposure methods, although inconsistencies were apparent between pollutants for some outcomes (Supplementary Material

Tables S3–S6). For example, SIMEX-corrected associations between STEAM modelled NO_2 estimates and all-cause dementia incidence (1.115 (0.980, 1.268)) increased the size of the uncorrected HR in the main model (1.062 (0.973, 1.160)), while the observed RCAL-corrected associations were even greater (1.179 (0.928, 1.496); Supplementary Material Table S6). However, associations between $PM_{2.5}$ estimates and all-cause dementia incidence provided more inconsistent patterns of effect estimate change after correction between exposure surrogates. The majority of the associations observed in these morbidity outcomes were not statistically significant, likely due to the low number of counts for disease incidence in our analysis.

4. Discussion

In the present study, several surrogate exposure estimates were compared assessing the effect of air pollution exposure on natural-cause mortality and COPD incidence both with/without measurement error correction methods applied. The effect estimates indicated increased risks with half of the surrogate measures, using the main adjustment model (Model 2), for all natural-cause mortality and with most measures for COPD incidence in the London-dwelling UK Biobank participants associated with long-term exposure mainly to NO_2, but also to $PM_{2.5}$. However, only the modelled estimates of ambient levels of NO_2 at the participants' residence were found to have statistically significant effects on COPD incidence, indicating an 8.7 % increase in risk per IQR increase in exposure. When SIMEX correction for measurement error was applied, all HRs increased (indicating a two-to four-fold increased risk of death or COPD incidence) and the associations for some surrogates of both NO_2 and $PM_{2.5}$ became statistically significant. A similar pattern was observed when the HRs were corrected using RCAL, but the confidence intervals of the corrected HRs were wider and the associations were not statistically significant in cases where uncorrected HRs were also not statistically significant. Health effect estimate correction also consistently resulted in widening 95 % CIs and therefore greater uncertainty, likely due to the small personal monitoring campaign (validation dataset) available. In some instances this observation was more pronounced when calculating standard errors via the Delta method for RCAL-corrected HRs, most likely due to the incorporation of the uncertainty in the calibration coefficients into the health effect estimation (Carroll et al., 2006; Richardson et al., 2025). The observed widening of confidence intervals is due to the implementation of the RCAL model (Keogh et al., 2020) and other previous analyses have shown confidence intervals to increase by more than 100 % after RCAL correction and use of the sandwich variance estimator (Hart et al., 2015b).

The increases in the HRs after measurement error correction identified in the present analysis are in close agreement with previous studies. Hart et al. (2015b) assessed the effects of a 10 $\mu g/m^3$ increase in $PM_{2.5}$ exposure on all-cause mortality and showed that the uncorrected HR increased from 1.13 (1.05, 1.22) to 1.18 (1.02, 1.36), and from 1.12 (1.05, 1.21) to 1.22 (1.02, 1.45) for modelled or nearest monitor measurement exposures, respectively, when RCAL was applied. Similar to the findings of the present study, the observed increase in HR magnitude was accompanied with greater uncertainty in effect estimates, evident from the increased standard errors and confidence interval width. Another study utilising the US Medicare cohort used stratified RCAL and assessed the impact of exposure measurement error, defined as the difference between modelled and measured ambient concentrations, on mortality effect estimates (Feng et al., 2023). The estimated measurement error bias resulted in underestimation of the true effect, but relatively small compared to our findings, i.e., up to 5.2 % towards the null, but this is probably due to the error definition which only included error in ambient exposure outside of residences and did not account for personal exposures (Wei et al., 2022; Katsouyanni and Evangelopoulos, 2022). Furthermore, simulation studies using London data have previously shown that for both NO_2 and $PM_{2.5}$, measurement error can bias the exposure-response associations by up to 98 % and 68 % towards the null, respectively, which is closely aligned with our findings (Samoli et al., 2020; Butland et al., 2020).

Interestingly, suggestive negative associations, i.e., protective effects of O_3 with both natural-cause mortality and COPD incidence (which in some cases reached the nominal level of statistical significance) changed direction, indicating adverse effects which were statistically significant after measurement error correction methods were applied. This finding, together with some evidence in the epidemiological literature that shows protective effects of O_3 on health (Hvidtfeldt et al., 2019; Liu et al., 2021), may suggest that measurement error bias in the health effect estimates of O_3 could be substantial and lead to false negative associations. However, this finding may be due to the high inverse correlation between NO_2 and O_3 often observed in measurements and models. The Pearson correlation coefficients for NO_2 and O_3 were −0.63 and −0.99 for nearest monitor and STEAM estimates, respectively. More personal exposure measurement campaigns for O_3 would provide further insight and could be used as validation samples for measurement error corrections in large epidemiological studies. Additionally, given the large number of associations tested, the potentially higher likelihood of observing statistically significant results due to multiple testing is a potential issue in our findings that should be noted.

We further assessed other morbidity outcomes, including incidence of MI, asthma, stroke and dementia but the analyses showed no effects (using the main adjustment model) of long-term exposure to NO_2 and $PM_{2.5}$, except for a suggestive increased risk for dementia related with increased NO_2 exposures (Supplementary Material Tables S3–S6). Similar to mortality and COPD incidence, SIMEX and RCAL corrections resulted in health effect estimates that moved away from the null, with two-to three-fold differences in some exposure-response associations. The fact that incidence was used and prevalent cases at baseline were excluded had an effect on the number of cases the present study was able to analyse and may have affected statistical power for some outcomes.

Most studies in the literature using UK Biobank data investigated morbidity outcomes but relatively few assessed air pollution effects on mortality. All such studies used UK-wide data and modelled air pollution exposure assessment already available in the UK Biobank database, such as the land-use regression models developed in the ESCAPE project (Eeftens et al., 2012; Beelen et al., 2013). Exposure variables in the present study were linked to the UK Biobank database as part of the MELONS project and it is the first time that a model taking time-activity data into account (LHEM-adjustment (Smith et al., 2016);) has been used in this dataset. Both exposure contrasts and residual confounding is expected to be different in the whole of the UK compared to the analysis provided here, which only included residents of London. Doiron et al. (2019) evaluated air pollution exposure, lung function and COPD in UK Biobank in a cross-sectional analysis and reported an increase in COPD prevalence associated with higher $PM_{2.5}$ and NO_2 exposures. However, the results were not directly comparable to ours, which assessed COPD incidence. Li et al. (2024) conducted an analysis on the risk of air pollution exposure with ischemic stroke incidence and reported statistically significant increased risks associated with both $PM_{2.5}$ and NO_2 exposures, in contrast with the findings presented here. For dementia incidence, Chen et al. (2023) observed HRs of 1.09 (1.06, 1.13) and 1.13 (1.09, 1.16) for IQR increases in $PM_{2.5}$ (2.3 $\mu g/m^3$) and NO_2 (about 20 $\mu g/m^3$), respectively, which are comparable to the results presented here.

Few publications have assessed the effects of air pollutants on UK-wide all-cause mortality, compared to London-wide natural-cause mortality (which was our main health outcome) using UK Biobank Data (Wang et al., 2022; Guyatt et al., 2024; Li et al., 2023). Wang et al. (2022) and Li et al. (2023) reported HRs per 10 $\mu g/m^3$ increase in pollutant concentrations, which is a large increment for $PM_{2.5}$ (when the IQR was reported as approximately 2 $\mu g/m^3$), but it is a reasonable contrast for NO_2 estimates. Guyatt et al. (2024) reported HRs per IQR of 9.7 $\mu g/m^3$ for NO_2 (not statistically significant) which are slightly higher

to those presented here (1.04 for never smokers and 1.05 for ever-smokers vs 1.03 in the present analysis per 14.54 μg/m^3 increase in STEAM estimates), but larger for $PM_{2.5}$. Interestingly, Wang et al. (2022) and Li et al. (2023) reported different HRs, perhaps due to different exposure modelling approaches, with Li et al. (2023) observing larger HRs. Both papers reported larger HRs compared to the present analysis. Converting the effect estimates in these studies to reflect the exposure increments used in the present analysis for STEAM estimates (IQRs of 14.54 and 1.72 μg/m^3 for NO_2 and $PM_{2.5}$, respectively), the HRs were 1.07 (1.01, 1.12) and 1.04 (1.01, 1.08) in Wang et al. (2022), as well as 1.18 (1.15, 1.21) and 1.11 (1.09, 1.14) in Li et al. (2023), for NO_2 and $PM_{2.5}$, respectively, compared to the uncorrected HRs of 1.03 (0.98, 1.07) and 1.00 (0.96, 1.04) reported in the present study.

The application of exposure measurement error correction methods in a survival analysis conducted on the UK Biobank cohort presented here offers several strengths over previous similar work. The size of the cohort presents a unique opportunity to assess both RCAL and SIMEX correction methods in a large (n = 61,797 participants) real-world setting. The use of multiple correction methods itself, which both provided increased HRs in comparison to uncorrected, is a strength of the present study although the two methods do differ in their underlying assumptions. Briefly, RCAL assumes that the "true" and surrogate exposures are linearly related and replaces surrogate values with the predicted "true". In comparison, SIMEX relies on finding a good fitting extrapolation equation and adds simulated error. These differences in approach may lead to differences in bias correction or variance estimation. SIMEX correction applies to the classical error component in the mixture, so it might not be the most appropriate method if the observed errors are predominantly of Berkson-like type. Additionally, if the errors were purely Berkson, a correction method would not be necessary, but our simulation work under review shows that even if the classical component is only 20 % of the mixture, RCAL/SIMEX corrections tend to provide more accurate health effect estimates.

Confounding adjustment was consistent across all outcomes and was decided *a priori*. The aim of the analyses presented here was to quantify exposure measurement error bias in epidemiological analysis. Although omitting an important confounding variable may have an impact on the observed quantified exposure-response relationships, the present study adjusted for an extensive set of potential confounders such as age, sex, ethnicity, smoking status and socioeconomic status, providing a robust analysis comparable to previous work from the ELAPSE project on morbidity outcomes such as COPD (Liu et al., 2021) and stroke (Wolf et al., 2021). Furthermore, we selected Model 2 as our main model as it accounted for multiple individual-level confounders. However, Model 1 comparisons with RCAL-correction were also included here to show the impact of confounding adjustment, as well as providing a comparison between corrected and uncorrected health effect estimates which share a common set of confounders across both the epidemiological and calibration models. We applied RCAL in Model 1 because the available set of confounders for the calibration model from the validation data was limited, and we wanted to assess the impact of applying RCAL in such instances. SIMEX does not require a calibration equation, so the correction was applied only on the final epidemiological model. To our knowledge, there is no previous publication that has compared the performance of these two methods on the correction of health effect estimation for exposure measurement error. However, limited previous work does exist comparing RCAL to multiple imputation correction, finding little difference in the associations between long-term $PM_{2.5}$ exposure and all-cause mortality in the US; likely due to the low variability observed in classical error (converse to that of the present study), as well as the definition of error given the lack of personal exposure in the validation dataset used (Josey et al., 2023).

Here, the utilisation of a validation dataset with long-term follow-up drawn from participants living in the same area (Greater London) as the study dataset and with similar surrogate exposure assessment, socioeconomic status and female/male balance, is a major advantage. The long follow-up of COPE participants allowed for the estimation of long-term personal exposures from outdoor sources only, an exposure metric not available in other panel studies in London. However, low correlations were observed between surrogate and "true" exposures, and we explored other panel studies to see if the correlations were similar or higher. In our recent paper (Zhang et al., 2025) we assessed four personal exposure measurement campaigns, i.e., COPE, one on schoolchildren (measuring personal exposure to $PM_{2.5}$), one on professional drivers (black carbon), one on healthy adults (black carbon) and partitioned total personal exposure into indoor- and outdoor-generated exposures (Zhang et al., 2025). We found low correlations between personal exposures and surrogate measures. Similar low correlations have been reported in previous studies in the US (Kioumourtzoglou et al., 2014). Thus, we considered COPE to be an appropriate personal monitoring campaign for this analysis. However, if the correlations between surrogate and "true" exposures were higher, it is likely that smaller differences between uncorrected and corrected health effect estimates would be observed (Samoli et al., 2020; Butland et al., 2020; Zeger et al., 2000).

Additionally, some potential limitations of using the COPE dataset do exist. For example, COPE individuals are older compared to UK Biobank participants who were about 56 years old on average at recruitment (2006-10). After approximately 15 years of follow-up they are of comparable age, however, COPE participants are individuals living with COPD and may therefore exhibit differing time-activity patterns which could affect personal exposure. This may impact the correction of surrogates towards a defined "true" measure of personal exposure in the validation dataset. Methodologies do exist in the measurement error literature to be implemented in cases where such differences are large (mainly borrowed from the survey statistics literature; e.g. (Oh et al., 2021; Barnatchez et al., 2024),), however, given the aforementioned similarities between COPE and UK Biobank we decided not to employ such methods and deemed the datasets similar enough. In the UK Biobank sample of Londoners used here, 2,081 (3.4 %) of participants were diagnosed with COPD throughout the course of follow-up (largest frequency in the outcomes under investigation), and thus, COPD was included in the main analysis of the present study.

A limitation of the present study is the lack of similar covariate information in the external validation dataset (COPE) to the data incorporated into the epidemiological analysis performed on the UK Biobank participants. Only Model 1 adjusted for the same covariates as the RCAL model, and this is the reason the present study also presented corrected HRs for this model in addition to the primary model (Model 2). Uneven adjustment between validation and study models may affect the error correction methods. However, good agreement was generally observed in the relative change in the estimates after RCAL adjustment between Model 1 and Model 2. In absolute values, the HRs of RCAL-corrected Model 1 and RCAL-corrected Model 2 were generally different, but this also held for the uncorrected HRs and is likely a result of the better confounding control. In addition, good agreement was observed between SIMEX-corrected Model 2 and RCAL-corrected Model 2 with considerable overlap in the confidence intervals between the two HRs. Additionally, a major strength of the use of COPE as a validation dataset is the long follow-up and compliance within the personal monitoring campaign itself. The COPE study followed up participants for a long period of time (up to six months) which is rare for such personal exposure studies due to cost and the inconvenience to participants (Moore et al., 2016). The inclusion of such a dataset potentially offers as strong an estimation of personal exposure as possible and therefore provides both RCAL and SIMEX correction methods with the best estimation of personal exposure to ambient sources, despite the relatively low correlations observed between "true" measured exposure and surrogate measures (e.g., 0.16 and 0.04 for NO_2 and $PM_{2.5}$ STEAM modelled estimates, respectively). Moreover, we assessed single exposure models and applied measurement error corrections separately for $PM_{2.5}$, NO_2 and O_3. However, measurement error bias may be more profound in

multi-pollutant models, where effect transfer can occur from the more poorly to the better measured exposures (Zeger et al., 2000) and measurement error correction could become more complex (e.g., error in one pollutant's exposure might distort another pollutant's effect estimate). Our future work will focus on quantifying measurement error bias in the effect estimates of long-term exposure to air pollution under a multi-pollutant model framework, given that the corrected HRs here pertain to single-pollutant effects and residual confounding by co-pollutants is still possible.

In conclusion, the present study showed that measurement error correction in the health effect estimates of long-term exposure to air pollution strengthens previous evidence that bias is generally towards the null, provided there is information on the error structures of the pollutants under investigation. Reported estimates not corrected for exposure measurement error may, in many instances, lead to underestimation of the magnitude of effects. We utilised a previous exposure measurement campaign of COPD patients with intensive follow-up and used personal exposures to outdoor-generated air pollution as the assumed "error-free" or "true" exposure. SIMEX- and RCAL-corrected HRs were substantially higher than the uncorrected "naïve" estimates, showing the importance of accounting for measurement error in our epidemiological analyses. These findings also have major implications in health impact assessment of air pollution exposure used to assess relevant policies. The use of corrected concentration-response functions may result in greater impacts which can justify more ambitious policies for improved air quality.

CRediT authorship contribution statement

Dimitris Evangelopoulos: Writing – original draft, Supervision, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Dylan Wood:** Writing – original draft, Resources, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Barbara K. Butland:** Writing – review & editing, Supervision, Investigation, Formal analysis, Conceptualization. **Benjamin Barratt:** Writing – review & editing, Supervision, Methodology, Funding acquisition, Conceptualization. **Hanbin Zhang:** Writing – review & editing, Methodology, Investigation, Formal analysis, Data curation. **Konstantina Dimakopoulou:** Writing – review & editing, Methodology, Investigation, Data curation. **Evangelia Samoli:** Writing – review & editing, Supervision, Methodology, Funding acquisition, Conceptualization. **Sean Beevers:** Writing – review & editing, Resources, Methodology, Investigation. **Heather Walton:** Writing – review & editing, Methodology, Funding acquisition, Conceptualization. **Joel Schwartz:** Writing – review & editing, Supervision, Methodology, Funding acquisition, Conceptualization. **Evangelos Evangelou:** Writing – review & editing, Resources, Funding acquisition, Conceptualization. **Klea Katsouyanni:** Writing – review & editing, Supervision, Methodology, Investigation, Funding acquisition, Formal analysis, Conceptualization.

Consent to participate

As a part of the COPE and UK Biobank recruitment processes, informed consent was obtained from all participants included in the studies.

Consent to publish

Not applicable.

Ethics approval

This study is an observational study that involves data collected on human participants. The COPE study was approved by the Independent Scientific Advisory Committee (ref 15052) and the Camden and Islington Research Ethics Committee (ref 14/LO/2216). Participants gave informed consent to participate in the study before taking part. The UK Biobank study was approved by the North West Multi-Centre Research Ethics Committee (ref 06/MRE09/65), and at recruitment all participants gave informed written consent to participate in UK Biobank and be followed up in accordance with the Declaration of Helsinki. This research has been conducted using the UK Biobank Resource under Application Number 22102.

Data sharing statement

This research involves personal data which cannot be made publicly available.

Funding

Health Effects Institute (HEI) funded under agreement #4974-RFA19-1/20-8-3.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: All co-authors reports financial support was provided by Health Effects Institute. Barbara K. Butland owns shares in Royal Dutch Shell and in Scottish and Southern Energy and her spouse is in receipt of a Shell pension. The other authors have no conflict of interest to declar

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 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. HEI retains a non-exclusive license to publish material from work funded by the HEI. DE, DW, BB, HW were part funded by the NIHR HPRU in Environmental Exposures and Health, a partnership between the UK Health Security Agency (UKHSA) and Imperial College London. DE was also funded by the MRC Centre for Environment and Health, Imperial College London. This research has been conducted using the UK Biobank Resource under Application Number 22102.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.envres.2025.122237.

Data availability

The authors do not have permission to share data.

References

Barnatchez, K., Nethery, R., Shepherd, B.E., Parmigiani, G., Josey, K.P., 2024. Flexible and Efficient Estimation of Causal Effects with Error-Prone Exposures: a Control Variates Approach for Measurement Error arXiv preprint arXiv:2410.12590.

Barratt, B., Zhang, H., Evangelopoulos, D., Walton, H., Katsouyanni, K., 2022. The use of snapshot personal exposure monitoring in estimating annual exposure to air pollutants for epidemiological studies. In: ISEE Conference Abstracts, vol.2022.

Beelen, R., Hoek, G., Vienneau, D., Eeftens, M., Dimakopoulou, K., Pedeli, X., et al., 2013. Development of no2 and nox 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.

Beevers, S.D., Kitwiroon, N., Williams, M.L., Carslaw, D.C., 2012. One way coupling of CMAQ and a road source dispersion model for fine scale air pollution predictions. Atmos. Environ. 59, 47–58.

Butland, B.K., Samoli, E., Atkinson, R.W., et al., 2020. Comparing the performance of air pollution models for nitrogen dioxide and ozone in the context of a multilevel epidemiological analysis. Environmental Epidemiology 4 (3).
Carroll, R.J., Ruppert, D., Stefanski, L.A., Crainiceanu, C.M., 2006. Measurement Error in Nonlinear Models: a Modern Perspective. Chapman and Hall/CRC.
Carslaw, D.C., Ropkins, K., 2012. Openair—An R package for air quality data analysis. Environ. Model. Software 27, 52–61.
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.
Chen, G.-C., Hukportie, D.N., Wan, Z., Li, F.-R., Wu, X.-B., 2023. The association between exposure to air pollution and dementia incidence: the modifying effect of smoking. J Gerontol A Biol Sci Med Sci 78 (12), 2309–2317, 2023.
Dimakopoulou, K., Samoli, E., Analitis, A., et al., 2022. Development and evaluation of spatio-temporal air pollution exposure models and their combinations in the greater London area, UK. Int. J. Environ. Res. Publ. Health 19 (9), 5401.
Doiron, D., de Hoogh, K., Probst-Hensch, N., et al., 2019. Air pollution, lung function and COPD: results from the population-based UK biobank study. Eur. Respir. J. 54, 1802140, 2019.
Eeftens, M., Tsai, M.-Y., Ampe, C., Anwander, B., Beelen, R., Bellander, T., et al., 2012. Spatial variation of PM2.5, PM10, PM2.5 absorbance and PM coarse concentrations between and within 20 European study areas and the relationship with NO2 – results of the ESCAPE project. Atmos. Environ. 62, 303–317. https://doi.org/10.1016/j.atmosenv.2012.08.038.
Evangelopoulos, D., Katsouyanni, K., Keogh, R.H., Samoli, E., Schwartz, J., Barratt, B., Zhang, H., Walton, H., 2020. PM2. 5 and NO2 exposure errors using proxy measures, including derived personal exposure from outdoor sources: a systematic review and meta-analysis. Environ. Int. 137, 105500.
Evangelopoulos, D., Zhang, H., Chatzidiakou, L., Walton, H., Katsouyanni, K., Jones, R. L., Quint, J.K., Barratt, B., 2024. Air pollution and respiratory health in patients with COPD: should we focus on indoor or outdoor sources? Thorax 79 (12), 1116–1123.
Feng, Y., Wei, Y., Coull, B.A., Schwartz, J.D., 2023. Measurement error correction for ambient PM2. 5 exposure using stratified regression calibration: effects on all-cause mortality. Environ. Res. 216, 114792.
Forastiere, F., Spadaro, J.V., Ancona, C., Andersen, Z.J., Cozzi, I., Gumy, S., Loncar, D., Mudu, P., Medina, S., Velasco, R.P., Walton, H., Zhang, J., Krzyzanowski, M., 2024. Choices of morbidity outcomes and concentration–response functions for health risk assessment of long-term exposure to air pollution. Environmental Epidemiology 8 (4), e314.
Guyatt, A.L., Cai, Y.S., Doiron, D., Tobin, M.D., Hansell, A.L., 2024. Air pollution, lung function and mortality: survival and mediation analyses in UK biobank. ERJ Open Research 10 (2).
Hansell, A., Kaewsri, P., Doiron, D., De Hoogh, K., 2024. Maximising research in environmental epidemiology using large cohorts such as UK Biobank–opportunities and pitfalls. In: ISEE Conference Abstracts, vol.2024.
Hart, Jaime E., et al., 2015a. Long-term ambient residential traffic–related exposures and measurement error–adjusted risk of incident lung cancer in the Netherlands cohort study on diet and cancer. Environmental health perspectives 123 (9), 860–866.
Hart, J.E., Liao, X., Hong, B., Puett, R.C., Yanosky, J.D., Suh, H., Kioumourtzoglou, M.A., Spiegelman, D., Laden, F., 2015b. The association of long-term exposure to PM 2.5 on all-cause mortality in the nurses' health study and the impact of measurement-error correction. Environmental health 14, 1–9.
Hoek, Gerard, Vienneau, Danielle, de Hoogh, Kees, 2024. Does residential address-based exposure assessment for outdoor air pollution lead to bias in epidemiological studies? Environmental Health 23 (1), 75.
Huangfu, P., Atkinson, R., 2020. Long-term exposure to NO2 and O3 and all-cause and respiratory mortality: a systematic review and meta-analysis. Environ. Int. 144, 105998.
Hvidtfeldt, U.A., Sørensen, M., Geels, C., Ketzel, M., Khan, J., Tjønneland, A., Overvad, K., Brandt, J., Raaschou-Nielsen, O., 2019. Long-term residential exposure to PM2. 5, PM10, Black carbon, NO2, and ozone and mortality in a Danish cohort. Environ. Int. 123, 265–272.
Josey, Kevin P., et al., 2023. Estimating a causal exposure response function with a continuous error-prone exposure: a study of fine particulate matter and all-cause mortality. J. Agric. Biol. Environ. Stat. 28 (1), 20–41.
Katsouyanni, K., Evangelopoulos, D., 2022. Invited perspective: impact of exposure measurement error on effect estimates—an important and neglected problem in air pollution epidemiology. Environmental Health Perspectives 130 (7), 071302.
Keogh, Ruth H., et al., 2020. STRATOS guidance document on measurement error and misclassification of variables in observational epidemiology: part 1—basic theory and simple methods of adjustment. Stat. Med. 39 (16), 2197–2231.
Kioumourtzoglou, Marianthi-Anna, et al., 2014. Exposure measurement error in PM 2.5 health effects studies: a pooled analysis of eight personal exposure validation studies. Environmental Health 13, 1–11.
Lederer, W., Küchenhoff, H., Lederer, M.W., Küchenhoff, M., 2019. Package 'Simex'. Package 'simex'.
Li, Z.-H., Wang, X.-M., Xiang, J.-X., et al., 2023. Associations of long-term joint exposure to various ambient air pollutants with all-cause and cause-specific mortality: evidence from a large population-based cohort study. Environ. Sci. Pollut. Control Ser. 1–11.
Li, P., Wang, Y., Tian, D., Liu, M., et al., 2024. Joint exposure to ambient air pollutants, genetic risk and ischemic stroke: a prospective analysis in UK biobank. Stroke 55 (3), 660–669, 2024.
Liu, S., Jørgensen, J.T., Ljungman, P., Pershagen, G., Bellander, T., Leander, K., Magnusson, P.K., Rizzuto, D., Hvidtfeldt, U.A., Raaschou-Nielsen, O., Wolf, K., 2021. Long-term exposure to low-level air pollution and incidence of chronic obstructive pulmonary disease: the ELAPSE project. Environ. Int. 146, 106267.
Moore, E., Chatzidiakou, L., Jones, R.L., et al., 2016. Linking e-health records, patient-reported symptoms and environmental exposure data to characterise and model COPD exacerbations: protocol for the COPE study. BMJ Open 6 (7).
Noble, S., McLennan, D., Noble, M., et al., 2019. The English Indices of Deprivation 2019. CLG Ministry of Housing, London.
Oh, E.J., Shepherd, B.E., Lumley, T., Shaw, P.A., 2021. Raking and regression calibration: methods to address bias from correlated covariate and time-to-event error. Stat. Med. 40 (3), 631–649.
Reeves, G.K., Cox, D.R., Darby, S.C., Whitley, E., 1998. Some aspects of measurement error in explanatory variables for continuous and binary regression models. Stat. Med. 17 (19), 2157–2177.
Richardson, Brian D., et al., 2025. Addressing confounding and continuous exposure measurement error using corrected score functions. Biometrics 81 (2), ujaf045.
Samoli, E., Butland, B.K., Rodopoulou, S., et al., 2020. The impact of measurement error in modeled ambient particles exposures on health effect estimates in multilevel analysis: a simulation study. Environmental Epidemiology 4 (3).
Shen, Y., de Hoogh, K., Schmitz, O., Clinton, N., Tuxen-Bettman, K., Brandt, J., Christensen, J.H., Frohn, L.M., Geels, C., Karssenberg, D., Vermeulen, R., 2024. Monthly average air pollution models using geographically weighted regression in Europe from 2000 to 2019. Sci. Total Environ. 918, 170550.
Smith, J.D., Mitsakou, C., Kitwiroon, N., et al., 2016. London hybrid exposure model: improving human exposure estimates to NO2 and PM2. 5 in an urban setting. Environmental science & technology 50 (21), 11760–11768.
Steinle, S., Reis, S., Sabel, C.E., 2013. Quantifying human exposure to air pollution—Moving from static monitoring to spatio-temporally resolved personal exposure assessment. Sci. Total Environ. 443, 184–193.
Strak, Maciej, et al., 2021. Long term exposure to low level air pollution and mortality in eight European cohorts within the ELAPSE project: pooled analysis. bmj 374.
Vodonos, A., Awad, Y.A., Schwartz, J., 2018. The concentration-response between long-term PM2.5 exposure and mortality; A meta-regression approach. Environ. Res. 166, 677–689.
Wang, M., Zhou, T., Song, Q., Ma, H., Hu, Y., Heianza, Y., Qi, L., 2022. Ambient air pollution, healthy diet and vegetable intakes, and mortality: a prospective UK biobank study. Int. J. Epidemiol. 51 (4), 1243–1253.
Wei, Y., Qiu, X., Yazdi, M.D., Shtein, A., Shi, L., Yang, J., Peralta, A.A., Coull, B.A., Schwartz, J.D., 2022. The impact of exposure measurement error on the estimated concentration-response relationship between long-term exposure to PM2.5 and mortality. Environmental health perspectives 130 (7), 77006.
Wolf, K., Hoffmann, B., Andersen, Z.J., et al., 2021. Long-term exposure to low-level air pollution and incidence of stroke and coronary heart disease: a pooled analysis of six European cohorts within the ELAPSE project. Lancet Planet. Health 5, e620–e632.
Wolf, Kathrin, et al., 2023. Comparison of traditional Cox regression and causal modeling to investigate the association between long-term air pollution exposure and natural-cause mortality within European cohorts. Environmental Pollution 327, 121515.
Wood, Dylan, et al., 2025. Personalised estimation of exposure to ambient air pollution and application in a longitudinal cohort analysis of cognitive function in London-dwelling older adults. J. Expo. Sci. Environ. Epidemiol. 1–8.
Zeger, S.L., Thomas, D., Dominici, F., Samet, J.M., Schwartz, J., Dockery, D., Cohen, A., 2000. Exposure measurement error in time-series studies of air pollution: concepts and consequences. Environmental health perspectives 108 (5), 419–426.
Zhang, H., Fan, Y., Han, Y., Yan, L., Zhou, B., Chen, W., Cai, Y., Chan, Q., Zhu, T., Kelly, F.J., Barratt, B., 2022. Partitioning indoor-generated and outdoor-generated PM2. 5 from real-time residential measurements in urban and peri-urban Beijing. Sci. Total Environ. 845, 157249.
Zhang, Hanbin, et al., 2025. Estimating exposure to pollutants generated from indoor and outdoor sources within vulnerable populations using personal air quality monitors: a London case study. Environ. Int. 198, 109431.