長崎大学
NAGASAKI UNIVERSITY

Investigation of long-term effects of $PM_{2.5}$ on mortality and cardiovascular hospitalizations on the UK Biobank cohort

Thesis submitted in accordance with the requirements for the joint degree of Doctor of Philosophy of the University of London and Nagasaki University

April 2025

Department of Public Health, Environments and Society
Faculty of Public Health and Policy
London School of Hygiene & Tropical Medicine

School of Tropical Medicine and Global Health
Nagasaki University

Funded by Nagasaki University "Doctoral Program for World-leading Innovative and Smart Education" for Global Health, "Global Health Elite Programme for Building a Healthier World" (WISE)

Research group affiliation: The Environment and Health Modelling Lab (LSHTM)

Contents

Declaration of own work

I, Jacopo Vanoli, confirm that the work presented in this thesis is my own. Where and if information has been derived from other sources, I confirm that this has been indicated in the thesis.

Acknowledgments

I want to express my heartfelt thanks to my supervisors, Antonio Gasparrini, Lina Madaniyazi and Chris Fook Sheng Ng, as each of them, in their way, has been encouraging, supportive and inspiring throughout this project. Even in proving circumstances, they consistently uplift my spirits. Thanks to colleagues in the Environmental Health Modelling (EHM) Lab, including Pierre Masselot, Malcolm Mistry, and Arturo de la Cruz-Libardi, for inspiring me throughout my whole PhD path. Thanks to Rochelle Schneider and Arturo de la Cruz-Libardi, who provided the exposure predictions that allowed me to perform all the project analyses. Lastly, I want to express my love and thanks to all my friends, especially those in Italy, London, and Nagasaki.

Chapter 1

Abstract

Ambient air pollution, especially fine particulate matter ($PM_{2.5}$), has been widely documented as a major global health concern linked to premature mortality and various chronic conditions. However, key limitations persist in the literature, including sparse individual-level data, imprecise historical exposure assessments, and uncertain confounding mechanisms. The UK Biobank, with its large sample size (>500,000 participants), comprehensive set of covariates, and detailed residential histories, provides an ideal platform to overcome these limitations. This thesis aims to quantify the chronic health effects of $PM_{2.5}$ using novel linkage frameworks, state-of-the-art epidemiological designs, and extensive confounder adjustment.

First, a linkage framework was developed to assign daily $PM_{2.5}$ exposures at a 1×1 km spatial resolution to each participant's residential address, accounting for changes in address over time. Second, time-to-event analyses were conducted using Cox Proportional Hazards models with time-varying exposures to estimate associations between chronic $PM_{2.5}$ exposure and all-

cause and cause-specific mortality, and cardiovascular hospital admissions. These models were further enhanced with distributed lag approaches to capture long-term lag structures. Finally, potential confounding pathways were examined using both individual-level (e.g., smoking, body mass index, physical activity) and area-level (e.g., neighborhood deprivation) data, facilitated by directed acyclic graph (DAG) analyses and stratified regression. The exposure-linkage process yielded high-resolution daily $PM_{2.5}$ estimates for the entire UK Biobank cohort, revealing substantial spatiotemporal heterogeneity. The output of the survival models demonstrated that participants exposed to higher levels of $PM_{2.5}$ experienced elevated risks for both all-cause and cardiovascular mortality, with effects persisting at relatively low pollutant concentrations. Analyses of hospital admissions underscored associations between $PM_{2.5}$ and an array of cardiovascular outcomes, including stroke subtypes and myocardial infarction. Distributed lag analyses indicated that 1-year and 3-year exposures exerted significant impacts, while 5-year averages sometimes revealed larger cumulative risks. Additional exploration of confounding established that area-level covariates, in particular recruitment centre and deprivation as well as individual behaviors each contributed to partially attenuating but not eliminating $PM_{2.5}$-related associations, highlighting a complex relationship between pollution levels and socio-behavioral factors.

These findings extend prior evidence by pinpointing key methodological advances in long-term $PM_{2.5}$ exposure assessment, demonstrating that temporally refined and spatially granular pollution models can uncover associations obscured in coarser exposures. The results also show that confounders at both the individual and neighborhood levels are relevant to unbiased risk estima-

tion, reaffirming the necessity of multi-scale data integration. Notably, the analyses highlight that meaningful health impacts may occur below current UK national regulatory thresholds, raising important public health questions regarding existing air quality standards.

This thesis demonstrates that integrating rich cohort data with high-resolution spatiotemporal air pollution estimates enables more accurate quantification of long-term $PM_{2.5}$-related health risks. Results emphasize that carefully addressing exposure assignment and multifactorial confounding can substantially refine effect estimates. The observed associations persist at relatively low exposure levels, suggesting a need for continued regulatory efforts. Beyond producing robust estimates for the UK context, the frameworks described here can be applied to other cohorts and environmental exposures, ultimately informing global strategies to reduce pollution-related disease burdens.

Chapter 2

Introduction

Ambient air pollution, and in particular fine particulate matter ($PM_{2.5}$), is a leading environmental risk factor, contributing substantially to early mortality and increased morbidity worldwide. While decades of epidemiological research have established consistent associations between $PM_{2.5}$ and adverse health outcomes — including cardiovascular disease, respiratory dysfunction, and, more recently, metabolic and neurological conditions — major uncertainties remain regarding the precise magnitude of these risks, the adequacy of existing regulatory thresholds in specific countries, and the mechanisms through which pollution drives chronic disease development.

A core challenge lies in the accurate reconstruction of long-term exposures at the individual level: many early cohort studies relied on sparse or centralized monitoring networks that provided only broad estimates of population exposure. Consequently, potential heterogeneity in pollutant concentrations (e.g., urban vs. rural areas) was sometimes masked, leading to measurement errors that could dilute or bias observed relationships. Equally critical is

the failure of many studies to address the temporal complexity of exposure, frequently collapsing multi-year $PM_{2.5}$ data into a single baseline measure that ignores meaningful fluctuations over time or potential lag structures in health effects.

Another persistent limitation is the inability to capture important sources of confounding. Although individual lifestyle factors—such as smoking, body mass index, and physical activity—are well-recognized in public health research, many large-scale administrative datasets do not collect or update such information at the granularity required for rigorous confounder control. Compounding this problem, residential mobility over the life course can significantly alter an individual's exposure profile. Failure to track address histories or to update exposure estimates accordingly may lead to underestimation or misclassification of risks, particularly for studies spanning multiple years or even decades. These issues become particularly salient when examining chronic, progressive health outcomes like cardiovascular disease, where the timing and duration of pollutant exposure may be especially crucial.

In the United Kingdom, a significant proportion of prior research on $PM_{2.5}$ has employed annual or single-time-point estimates from land-use regression models, with relatively coarse spatial coverage. While these methods have yielded valuable insights, they tend to overlook within-region variability and temporal changes in $PM_{2.5}$ concentrations, especially beyond major metropolitan areas. Moreover, UK-based studies often lack individual-level data on critical covariates or rely on area-level socio-economic proxies, potentially leaving important confounding pathways unexplored. In some cases, missing exposure data for regions like Scotland has forced investigators to exclude those participants, constraining the generalizability of their findings.

Against this backdrop, the need for comprehensive, high-resolution data that can account for residential mobility and capture nuanced pollution gradients has become increasingly evident—especially as pollution levels in the UK have undergone substantial secular declines over the past two decades.

The UK Biobank offers a powerful platform to address these limitations and deepen our understanding of air pollution's long-term health impacts. With more than half a million participants, detailed questionnaires, medical examinations, and linkage to national health registries, it provides a rare combination of size and depth. Crucially, the dataset includes full residential histories at fine spatial resolution, permitting the assignment of individualized $PM_{2.5}$ exposures that can be updated over time and aligned with specific follow-up windows. This level of detail enables sophisticated analyses that can disentangle short-, medium-, and long-term associations, while also assessing the role of potentially sensitive exposure periods. Moreover, the UK Biobank's robust measurements of individual- and area-level confounders—encompassing socio-economic status, lifestyle behaviors, comorbidities, and more—allow researchers to more accurately isolate the pollutant-related risks, shedding light on differential vulnerabilities across subgroups of the population.

Within this context, the goal of the current work is to leverage the UK Biobank's extensive repository of individual-level data and link it with high-resolution spatiotemporal $PM_{2.5}$ models. By reconstructing detailed exposure trajectories and applying advanced statistical methods such as time-varying survival analyses and distributed-lag approaches, this thesis aims to capture the complexity of the pollution–health relationship in the UK setting. Specifically, it explores whether the duration and timing of $PM_{2.5}$

exposure—rather than simple average concentrations—drive risks of premature mortality and cardiovascular events, and whether these risks remain detectable at levels below national and international regulatory thresholds. In doing so, the research addresses critical evidence gaps regarding the temporal dimensions of $PM_{2.5}$ exposure, residual confounding patterns, and the extent to which low-level pollution still contributes to the public health burden. By integrating detailed exposure assessment, thorough confounder control, and large-scale longitudinal data, the thesis attempts to improve current estimates of air pollution's health impacts that can inform subsequent epidemiological investigations throughout the UK and beyond.

Chapter 3

Aim and objectives

3.1 Aim

In this PhD project, I aimed to enhance our knowledge of the long-term effects of $PM_{2.5}$ on health outcomes through the investigation of the large prospective UK Biobank cohort, making use of high-resolution spatio-temporal exposure maps and state-of-the-art epidemiological methodologies. The conceptual framework that encapsulates this research is shown in figure 3.1.

3.2 Specific objectives

1. To link residential history records and high-resolution $PM_{2.5}$ maps to reconstruct individual-level daily exposure profiles for all the UKB participants.

2. To evaluate existing epidemiological designs and statistical methods aimed at studying health risks associated with chronic exposure to $PM_{2.5}$ with specific focus on approaches that facilitate the integration of de-

tailed individual-level data and the investigation of lagged temporal associations.

3. To apply methods from the previous objective to analyse the impact of long-term exposure to $PM_{2.5}$ on premature mortality and hospitalizations for cardio-respiratory outcomes using the large and rich database of the UK Biobank cohort.

4. To evaluate confounding mechanisms affecting cohort analyses on long-term exposure to air pollution from theoretical and practical perspectives.

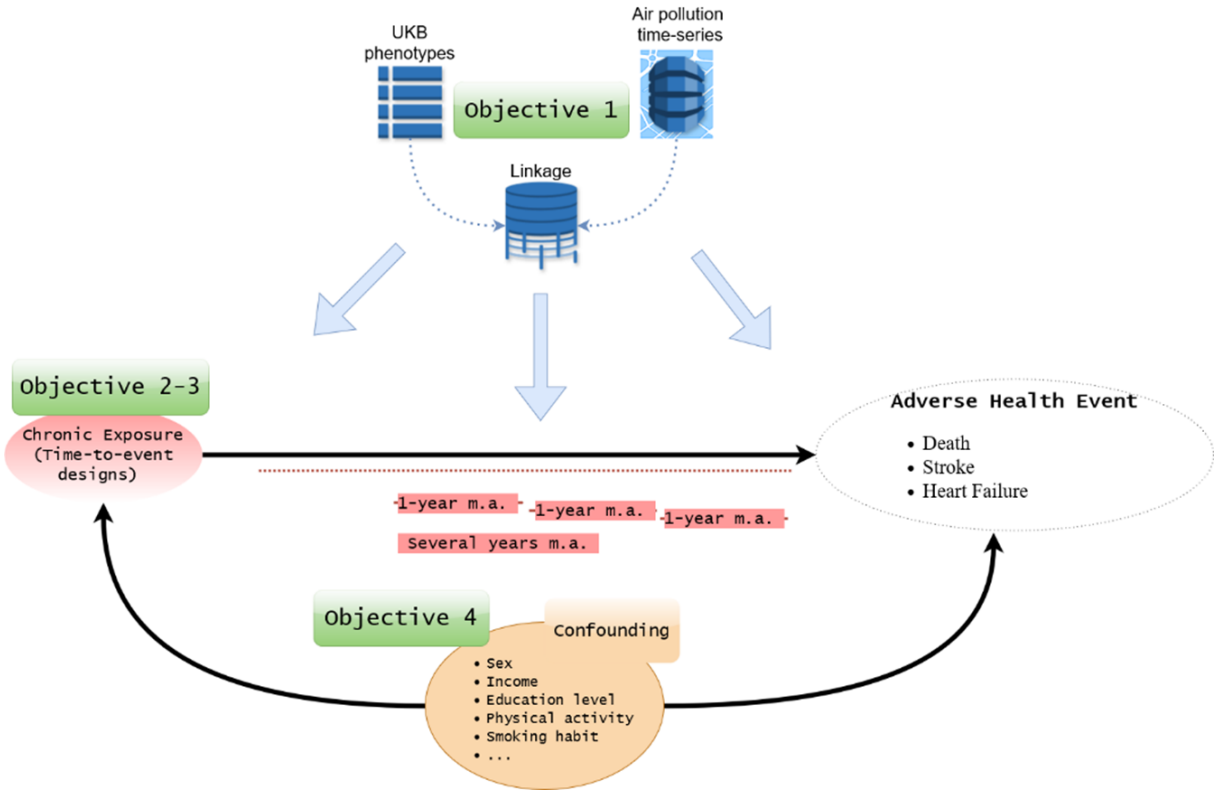


Figure 3.1: Conceptual framework of outdoor exposure to $PM_{2.5}$ and health outcomes.

Chapter 4

Definition of $PM_{2.5}$ and harmful effects

4.1 Definition of $PM_{2.5}$

Particulate matter (PM) is a significant pollutant comprising a mix of solid particles and liquids in the air[1]. It originates from various natural and anthropogenic sources, contributing to its ubiquitous presence in the atmosphere[2]. The majority of PM is produced by human activity, including combustion processes such as vehicle and industrial emissions, as well as house heating using fossil fuels [3, 4]. Additional important anthropogenic sources span from agriculture, such as livestock farming, and construction and demolition sites to biomass combustion for household activities[5, 6]. These various sources release particles that can be highly heterogeneous in their composition, mainly consisting of heavy metals, organic matter, soot and sulphates[2]. PM is typically defined based on the size of these particles:

PM_{10}, the larger inhalable particles, must have a diameter not larger than 10 micrometres, while $PM_{2.5}$, the fine particles, are limited to 2.5 micrometres and are considered the most dangerous for human health. Further down, ultrafine particles (UFP or $PM_{0.1}$) are also present and originate not only from combustion but also from gas-to-particle conversion processes. Although historically less studied due to limited data availability, recent advancements in measurement technology and modeling have made their investigation possible. The presence of PM in different sizes exacerbates air pollution along with other pollutants, such as NO_2 and O_3, which as standalone and synergically create a health burden to populations all over the planet[1, 7].

4.2 Physiological effects and cardiovascular diseases

PM, particularly $PM_{2.5}$, has long been part of the list of risk factors contributing to the global burden of disease and recognised as one the major causes of early death[8]. The scientific literature has produced evidence of its adverse effects on many health outcomes experimentally and epidemiologically. Investigation of biological mechanisms through which PM exerts its action is still ongoing[9], but specific physiological pathways have been identified[10]. One central mechanism involves increases in oxidative stress, where the fine particles induce the production of reactive oxygen species (ROS)[11]. This oxidative stress can lead to cellular and tissue damage, contributing to the onset and progression of various diseases. Metabolism dysregulation is another critical pathway[12]. PM has been linked to alterations in metabolic processes, which can result in disorders such as obesity, insulin resistance,

and Type 2 diabetes. Tissue inflammation is also an investigated response to PM exposure[13, 14]. The inhalation of fine particles can cause inflammation in the respiratory tract, which can then spread systemically, affecting organs and tissues throughout the body. Chronic inflammation is a known risk factor for many chronic diseases, including atherosclerosis[15]. Epigenomic changes represent another significant area of impact[16]. Exposure to PM can modify gene expression without altering the underlying DNA sequence. These epigenetic changes can impact cellular function and contribute to long-term health effects, including an increased risk of chronic diseases.

Although it is now well-established that $PM_{2.5}$ impacts health through multiple complex biological pathways, it is also clear that most of them are critical contributors to the development of cardiovascular diseases. Partly for this reason, cardiovascular conditions have become a central focus in both experimental and epidemiological research on the health effects of air pollution.

4.3 Impact of exposure windows on health

Assessment of when air pollution and environmental exposures, in general, exert their most detrimental impact can be formulated in two general categories: acute and chronic effects[17, 18]. They can be related to similar biological mechanisms described above but are characterised by different time frames in which pollutants increase the risk of a clinical event[19]. Regarding acute effects, the onset is mainly linked with ambient air pollution exposures experienced during the days preceding the event[17]. In this context, sudden changes in risk factors appear to cause the event directly. The sub-

jects with the event typically carry underlying vulnerabilities (such as old age or comorbidities) that make them prone to developing an acute clinical condition within hours or days of exposure to relatively high pollution levels. In this case, air pollution exposure acts as the trigger for the occurrence of the case[17]. On the other hand, the development of chronic conditions is associated with exposures cumulated over a long time frame, from years to decades. Therefore, long-term exposure to air pollutants does not trigger the occurrence of the acute event[19]. Still, instead, it favours its corresponding adverse physiological mechanisms, such as the development of atherosclerotic plaques, cardiovascular remodelling and changes in the epigenetic and immunologic makeup [16, 20–22]. Health risks related to long-term exposure are more of interest as they are linked with a substantially higher health burden[7]. However, due to more complex and expensive exposure modelling and epidemiological studies needed for the analysis of chronic effects, short-term investigations were prominent in the historical literature, while investigations on long-term risks were mainly concerning analyses and re-analyses of a few cohorts[23–25]. More recently, there has been a wealth of published investigations, mainly focusing on the relationship between chronic exposure to PM and mortality outcomes and cardio-respiratory hospital admissions [26–28]. This increase in the number of long-term studies was mainly due to advancements in exposure assessment methods and the availability of data from new cohorts[29]. In recent years, epidemiological investigations found long-term exposure to air pollutants to be associated with increased risk for several health outcomes, including respiratory diseases[30]. Mainly chronic obstructive pulmonary disease (COPD) and exacerbation of asthma[31], and cardiovascular events, such as myocardial infarction (MI)

and stroke[26]. Moreover, it has been suggested that air pollution mediates the occurrence of diabetes[32] and lung cancer[33] And there is growing evidence of effects on neurological disorders and cognitive function[34, 35]. In this thesis, I focused on the long-term effects of fine particulate matter ($PM_{2.5}$) on health outcomes, specifically premature mortality and hospital admissions for cardiovascular events.

The choice of the outcomes was based on several factors: first, as outlined in the previous chapter these outcomes have strong biological plausibility. Second, both mortality and cardiovascular diseases burden hold a prominent significance in the public health context. Finally, despite the existence of several studies on these topics, in the UK the literature is scarce.

Chapter 5

Epidemiological literature

5.1 Cohort studies

In the last three decades, a large body of epidemiological investigations focused on the chronic effects of $PM_{2.5}$ on health outcomes, particularly premature mortality. Until 2010, the great majority of the investigations took place in the US, with analyses and re-analyses of the American Six City cohorts[24, 36] and the American Cancer Society Cancer Prevention Study-II (CPS-II)[23, 37]. These studies detected a strong and consistent increased mortality risk due to fine particulate matter and effectively influenced policy-makers to target reduced particulate matter emissions by lowering annual air quality guideline values (the 2005 WHO limit was 10 $\mu g/m^3$, and later, the US EPA limit was 12 $\mu g/m^3$) and implementing practical interventions. In the following years, motivated by American studies, the European ESCAPE

project investigated long-term effects on mortality and hospitalisations across several countries[38, 39]. The ambitious project included cohorts from many countries, extensively controlled for confounding and investigated several outcomes. However, the lack of power from the relatively short follow-ups and the use of small cohorts implied non-significant results and criticisms[40]. Other attempts to investigate the long-term effects of $PM_{2.5}$ in Europe were rare[41–43]. Notably, only one English study[44] investigated all-cause and specific-cause mortality, finding weak associations, particularly for cardiovascular mortality, after controlling for confounders. Consequently, since 2016, the Health Effects Institute (HEI) funded three major studies in Europe[45], the US[46], and Canada[47] to investigate the effects of even lower levels of $PM_{2.5}$. Briefly, across five years, US, Canadian and European cohort data, including almost 100 million subjects, were used to explore the effects of pollutants, especially $PM_{2.5}$, below the previously set threshold. In North America, the Medicare[48], the Canadian Census Health and Environment Cohort (CanCHEC)[49], and its sub-sample, the Community Health Survey Cohort (CCHS)[50], were utilised in large-scale investigations. These investigations involved large samples and, therefore, were well-powered to investigate the effects of air pollutants on health outcomes. On the other hand, in Europe, several investigations within the ELAPSE[51–53] project included smaller national cohorts that were analyzed separately and their estimates meta-analysed using harmonised modelling approaches, similarly to ESCAPE[38, 41, 54]. In addition, ELAPSE investigated nationwide cohorts from Belgium, Denmark, England, Netherlands, Norway, and Switzerland, as well as one citywide cohort in Rome, Italy[55, 56]. Finally, the high heterogeneity in characteristics among the cohorts from different countries

prompted a sub-project to conduct investigations on a harmonised database to increase their comparability, finding relatively similar health effects. A detailed overview of the entire analysis is reported in a recent publication[57]. Moreover, the results of more than 20 years of literature were summarised in two reviews[27, 28]. These produced meta-analytical hazard ratios (HRs) ranging between 1.08 and 1.09 (by 10 $\mu g/m^3$ increase in exposure) for non-accidental mortality with confidence intervals from 1.06 (lower band) up to 1.11 (upper band). Eventually, these motivated the WHO to further reduce the annual limit to 5 $\mu g/m^3$.

In general, the feasibility of accurate long-term air pollution studies on health outcomes highly depends on the cohort data available to the investigators. The literature includes many different cohort data, mainly described above, which can be broadly split into two categories linked to their characteristics: administrative[55] and traditional[51]. These two types mainly differ in terms of the original purpose of the data. While administrative databases contain health data collected by governmental organisations for non-statistical purposes in origin, traditional cohorts are created explicitly for research purposes[27, 28]. In practical terms, this difference means that administrative data typically contain minimal personal information on the participants, as this is not the primary interest for data collection. Moreover, regarding research purposes, administrative databases typically require strict privacy-secure procedures for access and use, as initially, the individuals did not provide consent to analyse their data for research purposes. These two main differences reflect on the research that can conducted using them. Typically, data analysis of traditional cohorts is more flexible due to the wider set of data available, particularly on personal characteristics and lifestyles. Finally,

their size is an essential difference between these two types. Administrative cohorts, extracted from government databases, can be extremely large, including sometimes dozens of millions of subjects[48, 49]. In contrast, traditional cohorts are generally very small compared to their counterparts[51, 52], and therefore, they may lack statistical power to find significant associations with the long-term effects of environmental exposures.

5.2 Exposure assessment and linkages

In epidemiological air pollution studies, it is crucial to know how the exposure is assessed and assigned to the cohort investigated to identify potential design issues and interpret correctly the results of a study[58]. Exposure assessment is a process where researchers estimate an environmental exposure, such as $PM_{2.5}$, at varying degrees of temporal and spatial resolution. There exists a multitude of statistical model strategies to predict exposure measurements. The simplest method, often applied in the early studies[24, 37], was to assign the values obtained from the closest ground-monitoring station. However, in the last decades, novel methodologies have been developed to obtain exposure measurements with improved accuracy[29, 59]. Land use regression (LUR) methods combine land-use information with demographic and ancillary data, offering good performance at reduced geographic scales[60, 61]. However, they are less effective for modelling pollutant levels across large areas and in time[62]. These models have been mainly used to predict exposure for European epidemiological studies[45]. In contrast, emission-dispersion models (EDMs) simulate the physical transport of pollutants emitted from known sources[63]. They are suitable for modelling

air pollution on large geographical scales but require high-powered computing to simulate meteorological and chemical phenomena[64]. Satellite data have also been used to estimate ground pollution levels. Still, despite their ubiquity, satellite data alone cannot accurately capture ground-level concentrations and are affected by large measurement gaps due to cloud cover and sun-earth surface reflectance[58]. More recently, methods based on machine learning (ML) algorithms have provided powerful tools to predict ambient exposure at high accuracy[29, 65]. These methods merge multiple data types from various sources described above (monitors, land-use variables, satellites and EDMs) into complex multi-stage statistical modelling frameworks that, in most cases, outperform traditional techniques by providing more accurate estimates of pollutants, especially at fine spatial and temporal resolutions[59]. These model frameworks have been used in the last decade to predict daily air pollution levels at 1x1 km, mainly across North America[59, 66]. However, analogous models have also been applied elsewhere[67–69]. Subsequently, exposure assignment is a process designed to link the exposure estimate with the population in which we are interested in assessing health effects. Accurate exposure assignment is essential for epidemiological studies to quantify the impact of environmental factors on health outcomes correctly. The choice between the type of exposure to assign, measured or predicted, depends on the study context. Differently from the modelled predictions described before, exposure measurements involves the direct collection of the pollutant's data, either by fixed home monitors or mobile devices. This type of real-time monitoring campaigns are extremely financially and logistically costly, even for studies with a few hundred participants and relatively short follow-up periods (one months to a year). Therefore, monitoring individual

exposure is preferred when high accuracy and specificity are necessary. An example is represented by studies on vulnerable populations in specific occupational settings, where personal monitoring devices offer the most accurate measure of actual exposure. However, these methods are not a viable option for retrospective long-term effect studies that require several years of historical exposure data to align with the cohorts' follow-up periods. Nonetheless, direct measurements on a subset of individuals can be beneficial for evaluating and improving the accuracy of air pollution models and interpolation techniques. Finally, real-time exposure monitoring is crucial when the aim of a study is to assess the effect of air pollution on subclinical health outcomes, such as temporary increases in blood pressure, that cannot be captured with predictions performed in subsequent times. Because the aim of my doctoral project is the assess long-term effects of $PM_{2.5}$ I focused my attention on modelled exposure predictions and on how those are assigned to study participants. Commonly, the investigators assign predicted exposure at the personal residential address, assuming that a subject will spend most of the time in the surroundings of their residence or not too far from it[70]. This is the primary method applied in epidemiology when residential addresses are available. However, this is not always possible. For example, personal residential addresses are typically unavailable when using administrative cohort data due to privacy constraints[48]. Therefore, the investigators must assign the data to the first territorial unit for which the data are available. This may span dozens, if not hundreds of kilometres to the detriment of the accuracy of the assignment. In this thesis, I described and implemented a framework that expands the toolkit for epidemiologists for linking modelled environmental exposure to cohort data with fine spatial and temporal resolu-

tion. I applied this framework to assign daily $PM_{2.5}$ estimates predicted over a 1x1km grid (see Appendix) to the residential addresses of the UK Biobank participants.

5.3 Study design

Analysing event outcomes over time necessitates using a time-to-event design involving statistical survival models. The Cox Proportional Hazard (Cox PH) model is a technique commonly applied for censored data that can be used to analyse the survival time of study subjects by defining risk sets that include both the cases and all the non-cases present in the study at each event time[71]. It is prominently utilised in the literature to evaluate the long-term impacts of air pollutants, and there are several reasons for its widespread adoption. First, historically, the Cox model was implemented in seminal studies, such as the Harvard Six Cities. These pioneering applications yielded significant and consistent results, establishing the model's credibility and leading to its continued use in subsequent research[72, 73]. Second, the Cox model offers the advantage of controlling for multiple covariates by simply adjusting for them in the model, eliminating the need for complete data stratification. This feature enables researchers to simultaneously account for several potential confounders that might otherwise bias the exposure-response relationship. Finally, the Cox PH model allows extensions for the inclusion of time-varying covariates in a structured and controlled manner, as opposed to time-fixed exposure. This occurs thanks to a redefinition of this model for counting processes[74], which involves the specification of a dataset with multiple observations per subject, each representing an interval

time with the follow-up. Associated with this data, an additional indicator variable specifies if the subject-interval observation includes the event. This extension allows us to easily incorporate time-dependent exposure by assigning different exposure levels to each split period. Interestingly, this extension of the Cox model potentially will enable researchers to use complex methodologies to investigate the temporality of the effects, such as the distributed lag models (DLM), using recently developed methods [75]. These models, initially developed for short-term design settings, have rarely been applied in modelling long-term effects[76]. However, time-varying exposure coupled with DLM tools could enable investigators to identify the most relevant exposure windows over mid- to long-term time frames, namely from months to several years. This could shed light on the timing of the pathogenesis of different chronic and acute clinical conditions. The Cox model requires the definition of a specific time axis, which represents the timescale chosen to stratify the data into the risk sets for each event. Typically, the choice is between the age of participants and calendar year, although other axes, such as time since entry, have been used in other contexts. The selection may depend on the type of exposure used in the study. If the study uses a time-fixed exposure, then there is no need to control for calendar year because the exposure will not vary in time, and therefore, using age as a time axis along stratification for the year of enrolment will be enough to account for temporal confounding[51, 55]. Conversely, the calendar year can be more suitable as the main time axis when the exposure used is also time-dependent. Each risk set, as described above, corresponds to a specific time interval over which exposure summaries need to be matched. Using the correct time axis is essential when choosing a long-term study design. Additionally, the continuous

increase in the size of the data used in the analyses has encouraged the use of always more sophisticated analysis tools. From the use of complex hardware architectures (e.g., computer clusters) to the use of different modelling techniques, such as the equivalence with the Poisson model[46, 77] or the linear probability model[78] to reduce the magnitude of the data. In this thesis, I used the time-varying exposure metrics with high spatiotemporal resolution and DLM models to study exposure-response associations between $PM_{2.5}$ and mortality and cardiovascular inpatient hospitalisations[75, 76]. Specifically, I conducted studies analysing differences in the association when exposure windows of varying lengths were applied (for instance, 1-year and 5-year time-dependent windows). Moreover, for mortality outcomes, I used DLM to investigate the decomposition of the effect into single-lagged years. Finally, during my project, I also considered an effective alternative to the Cox PH model by exploring the risk-set sampling techniques (see 14).

5.4 Confounding in long-term studies

In epidemiology, a covariate is a variable associated with a study's outcome. Confounders are a specific type of covariate that affects both the exposure and the outcome but are not influenced by them[79]. These variables are crucial when studying the exposure-outcome relationship, as their omission can lead to biased estimates when evaluating associations[80]. Traditional confounders mainly include information on study subjects regarding anthropometric measures, socio-economic status, lifestyles and behaviours (i.e. individual-level covariates)[80, 81]. However, other types of confounding also exist in the form of contextual (i.e. area-level) variables and temporal

factors[82, 83]. The former can include environmental, social, economic and cultural factors characterising the surroundings of the subject's residence in the study. Differently from individual-level, contextual confounders operate at neighbourhood or community levels; therefore, instead of being related to a single subject, they are related to groups of subjects residing in the same area. On the other hand, temporal confounders are timescales (not classical variables) across which exposure and outcomes exhibited specific patterns that, if interdependent, may lead to biases. An important example of temporal confounding is represented by secular trends in $PM_{2.5}$ exposure in terms of magnitude (i.e. exposure levels) and composition[84] and unrelated changes in underlying baseline risks, which can confound the pollutant's relationship with health outcomes over time[83]. Owing to their intuitive application, timescales are usually accurately considered and effectively incorporated in the literature[48, 55]. In air pollution epidemiology, a main distinction occurs between individual- and area-level definitions of socio-economic status (SES)[82]. With the first, researchers consider all the personal information, including income, individual education, and employment, measured at a specific point in time (typically at the beginning of the follow-up), while the second includes the same information at the area level (e.g. % of subjects with a university degree) as well other measures that can be considered proxies of "area-level wealth", such as indexes of deprivation. This is an important difference because while individual-level factors have been more rarely investigated, the role of area-level covariates is now well-recognized[80, 85]. This is likely due to the analogous causes that might influence both neighbourhood/community resources and environmental factors, leading to a spatial correlation among them[85]. These patterns have

been thoroughly investigated for socio-economic status (SES) covariates[82]. Essential tools to identify relevant confounders are directed acyclic graphs (DAGs)[86]. These help visualise and analyse the web of causal relationships among variables by graphically representing them. In the past, they were rarely applied[87–89], but the use of DAGs in air pollution epidemiology has grown in the last decade. For instance, a comprehensive study conducted in Canada[49] exemplifies the use of DAGs in conjunction with traditional confounders selection to pinpoint the minimal set of critical covariates to include as relevant confounders in their analyses. In their findings, the authors detected a variation in the results between the traditional and the DAG model. In general, this implies that when investigating the same outcome within the same cohort, the selection of confounders can vary based on the investigators' choices, potentially leading to different results[81]. This variability motivates the importance of the initial assumptions and decisions made during the design of the study, as they can significantly influence the findings[90]. In this PhD research, I have extensively used covariates at both area and individual levels, analysing their potential confounding effects and the related impact on the estimates. In particular, I will use DAGs and regression models to analyse their potential role in the analyses.

Chapter 6

UK Biobank

6.1 Overview

The UKB Study is a comprehensive prospective cohort study designed to explore the genetic and lifestyle factors influencing various diseases common in middle and later life. To recruit participants, UKB researchers sent postal invitations to 9,238,453 individuals aged 40-69, who were registered with the UK's National Health Service and resided within approximately 25 miles (40 km) of one of the 22 assessment centres spread across England, Wales, and Scotland. Between 2006 and 2010, a total of 503,317 participants agreed to join the study and attended an assessment centre, with a participation rate of 5.45%[91, 92]. The assessment visit comprised electronic signed consent; a self-completed touch-screen questionnaire regarding lifestyles, behaviours and personal characteristics; a brief computer-assisted interview; physical and functional measures; and blood, urine, and saliva collection. UK Biobank also routinely conducts additional phenotyping assessments in participants' subsets, and genotypes and genomic data are avail-

able. The cohort has been tracked directly through follow-up assessments involving subsets of subjects and indirectly through routine linkage with various administrative health databases, including mortality, hospital episode statistics, cancer screening, and primary care visits. Notably, at recruitment, participants provided consent to use their residential address and follow-up information on the residential histories was provided either through the participants' notice or through linkage with the NHS system. Specific details regarding the UKB database can be found on the showcase website (https://biobank.ndph.ox.ac.uk/showcase/).

6.2 Historical particulate assessment and health impact studies

Air pollution data originally available in the UKB database were produced by European LUR models. Specifically, $PM_{2.5}$ estimates for the year 2010 were modelled for each participant's address using a LUR model developed as part of the ESCAPE project. The LUR model is based on field monitoring campaigns done in London between 26 January 2010 -18 January 2011, and air pollution estimates are representative of the year 2010 only. Traffic variables were calculated within a geographic information system (GIS) during LUR. ESCAPE estimates for particulates are extrapolated and deemed valid up to 400km from the London monitoring area, although it is unclear how reliable the estimates are outside this area. All addresses which are more than 400km away from Greater London are not assigned PM_{10}, $PM_{2.5}$, $PM_{2.5}$ absorbance and PM coarse concentrations and have missing data instead (number of missing records: 33,935). R^2 validation for UK in Manchester is equal to

21% and in the London-Oxford area equal to 77%(60).

Further details about NO_2 and NO_x estimates can be found in [61], while details on estimates of PM_{10}, $PM_{2.5}$, $PM_{2.5}$ absorbance and PM coarse are reported in [60]. Several studies were conducted on the associations between fine particulate health outcomes using the exposure estimates summarised in the above paragraph, and various outcomes were analysed. Cross-sectional studies[31, 93] examined the effects of pollution on the prevalence of respiratory diseases and lung function. A longitudinal study[94]investigated the impact on the incidence of cardiovascular hospital admissions, finding positive associations with $PM_{2.5}$. Also, adverse cardiac phenotypes were correlated with air pollution exposure[95]. Investigations on cognitive performance[96] and brain volume[97] suggest that exposure might harm mental cognition and reduce certain brain lobes' size, respectively. Eye health was also shown to be affected by air pollution in two studies on glaucoma prevalence[98] and adverse retinal structural features[99]. This is only a brief list of the publications generated using the current UKB exposure data, but it illustrates the research potential of this database.

Chapter 7

Limitations of the current literature

As described in the previous sections, there is a wealth of literature on health risks associated with exposure to air pollution and $PM_{2.5}$. However, limitations and gaps are still present, and new studies benefitting from novel methods and data resources can provide insights to improve our knowledge on this topic.

7.1 Exposure assessment and time-varying exposures

From the perspective of exposure assessment, the literature differs significantly across countries. In North America, the last decade has seen the development of exposure models with high spatial and temporal resolution across large areas. For example, US researchers have developed and ap-

plied hybrid ML models to predict daily pollutant measurements across a 1x1 km grid for all the contiguous states[59]. At the same time, Canadian studies have used complex predictions from models incorporating ground monitors and satellite data to predict annual air pollution measurements back to 1981[66]. These efforts were crucial for long-term studies, allowing investigators to conduct survival analyses for massive sample sizes (see Cohort studies section) using detailed time-dependent exposure summaries. In contrast, European studies (mainly ESCAPE and ELAPSE projects) have relied on time-fixed LUR models to investigate long-term health effects[60, 100]. These models produced exposure estimates at a very high resolution (100x100 m) for multiple pollutants across Europe in 2010. Very recently, LUR Europe-wide predictions have been replaced by back-extrapolated exposure series with the aim of obtaining updated associations[55, 57] more comparable with North American studies. The LUR predictions used in the European studies have also been linked to the UK Biobank database (see 6.2) and have been used extensively in the last years to assess health associations. Although the characteristics of time-fixed exposure metrics make them easily applicable in standard survival studies, it is inappropriate for investigations with more complex designs, such as time-varying survival models[74]. Also, time-fixed metrics do not reflect the actual individual exposure profiles as time passes[18]. It has been shown that using time-fixed exposure instead of time-dependent leads to downward biased associations, mainly if underlying trends occur[101]. This is especially the case in studies on the long-term effects of air pollution, as in the last decade, concentration levels have decreased substantially, especially in Western countries[84]. Even assuming a constant spatial exposure distribution across time, secular changes in partic-

ulate matter can impact the health associations' estimates due to changes in the PM composition[28] and unpredictable and unknown processes. Finally, time-fixed exposure data cannot be used to investigate medium- and short-term effects when the time granularity required is in the order of daily, weekly or monthly averages. In addition, the LUR Europe-wide predictions include completely missing data for Scotland, leading to the exclusion of 10% of the UK Biobank cohort when performing statistical analyses.

7.2 Exposure assignment

Air pollution predictions are typically assigned at the residential level, but the spatial resolution of the linkage changes widely across studies. In the Medicare database, residential data are only available at the ZIP code level, with areas that can span from a few hundred meters to thousands of square miles[59, 83, 102]. Canadian CanCHEC-based studies are characterised by a similar postcode linkage despite the corresponding areas being smaller (1-5 km) compared to the US definitions[49, 50, 103]. In these conditions, exposure assignment can only be made by averaging the exposure across the entire postcode area and linking the corresponding value to the Medicare subject. However, exposure summaries across several dozens of miles are unlikely to represent the exposure corresponding to a specific address, and previous studies have shown a decrease in the estimated health risks when the spatial buffer increased[104–106]. On the other hand, studies using European cohorts (ESCAPE AND ELAPSE projects) performed the linkage directly with the residential addresses, granting a more accurate spatial description of the subjects' exposure levels, although the temporal resolution was limited

to a one-year annual average[38, 51, 55].

7.3 Windows of exposure

Another significant issue pertains to the timing of exposure in relation to its effects. The literature on identifying the most relevant time windows for exposure is limited, with most research focusing on acute effects and leaving a significant gap in our understanding of the temporal dependencies associated with long-term exposures. Nevertheless, a few studies have explored this area. For instance, an historical work [89] examined the impact of $PM_{2.5}$ on mortality risk across different annual exposure lags, finding no significant association beyond the year immediately preceding the event. Two other studies that investigated how the length of the exposure window influenced associations found mixed results regarding mortality and cardiovascular hospitalisations. In Lefler et al., 2019 [105], the authors analysed a sizeable U.S.-based cohort and found consistent mortality associations for 2-year and 5-year exposure averages. In contrast, Crouse and colleagues [104] observed more substantial effects with longer exposure windows in a Canadian census-based study. Given the scarcity of research in this area, further studies are needed to understand these temporal dynamics better. Distributed Lag Models (DLMs) offer a promising approach for investigating the impact of different exposure windows and they are already applicable in cohort analyses [56] , making them well-suited for long-term air pollution studies.

7.4 Control for confounding

Up until 2020, the findings from 30 years of research in Western countries were summarised in two meta-analyses, which examined the effects of chronic pollution exposure on all-cause and cause-specific mortality[27, 28]. These analyses identified a few controversies relevant to this thesis, particularly regarding diverse confounders' adjustment strategies applied in the literature. One key issue is that, due to the administrative nature of multiple cohorts (including Medicare and CanCHEC), only a limited number of studies[23, 44, 51, 107] controlled for individual confounders, such as personal characteristics and lifestyle factors. European studies with traditional cohorts predominantly involved small, traditional cohorts, which lacked the statistical power to produce consistent results and were challenging to harmonise regarding the confounder definitions[40]. In contrast, large cohorts in North America, such as the ACS[23], NIH-AARP[108] and CCHS[47, 50] allowed for more control over individual confounders. However, many of these studies focused primarily on smoking status and BMI, often neglecting other potential behavioural covariates like physical activity and drinking patterns[23, 108]. It's important to consider these differences, as the role of individual covariates as confounders has been debated in the literature [41, 86]. For example, some historical studies have found that behavioural confounders, aside from smoking, did not significantly affect confounding [80, 109], while others have shown that including these behaviours altered the observed associations [110]. Although individual characteristics beyond smoking status and BMI are unlikely to have a strong confounding effect, it is essential to define consistent sets of confounders when investigating the long-term effects of air pollution

for multiple reasons. First, ensuring consistency across studies is crucial, given that heterogeneity in study characteristics and exposure definitions complicates direct interpretation and the drawing of consistent conclusions. While the literature has traditionally been heterogeneous, recent efforts have been made to conduct more harmonised studies[57]. Furthermore, improving study validity within the same cohort is particularly important, especially in the context of the UK Biobank. The accessibility and wealth of covariates in the UK Biobank can lead researchers to apply varied confounding adjustments without identifying the most relevant and important covariates. This flexibility may result in cherry-picking confounders or outcomes based on the significance and relevance of the results.

7.5 Addressing gaps and limitations

In this section, I concisely anticipate the novel aspects of my PhD project in light of the objectives and limitations of the literature.

In this work I implemented a high-resolution exposure linkage framework that integrates daily $PM_{2.5}$ concentration estimates with individuals' full residential histories, documenting every step of the process. This dynamic approach addresses a key limitation of past European studies, which mostly used exposure metrics (e.g. one-time or averaged values) that failed to capture changes in location or temporal pollution variation. By reconstructing continuous, day-by-day exposure profiles for each participant, the framework overcomes exposure misclassification inherent in previous methods. Moreover, due to the relatively fine resolution of the exposure, this linkage enables a researchers to define flexible exposure summaries tailored to the study design needs.

For the epidemiological investigations, in this project I analyzed and tested state-of-the-art survival models for time-dependent exposure. This corrects potential biases that arise from treating long-term exposure as time-fixed; instead, each person's $PM_{2.5}$ exposure was updated throughout the follow-up, aligning risk estimates with actual exposure history. Using the above methodologies, I produced new evidence on long-term $PM_{2.5}$ effects on mortality to extend the existing knowledge, particularly in the UK context. This research will also provide one of the most comprehensive long-term assessment of fine particulate matter's impact on cardiovascular hospitalizations in the UK and Europe, yielding insights that address gaps in prior literature which lacked such detail. Additionally, I implemented DLMs for chronic exposure effects to identify critical windows of exposure. This approach is a methodological improvement that may uncover relevant temporal patterns (e.g. whether recent vs. earlier exposures drive risk) that previous assignment could not detect.

Finally, during the whole project I employed rigorous confounder control through the use of DAGs and thoughtful covariate selection. By doing so, this thesis will offer practical guidance for future research with the UK Biobank: for example, always include regional controls for spatial confounding to address residual confounding. This level of guidance is an advancement over conventional practice, where often a "standard" set of confounders is used without exploration of its adequacy.

The work collectively enhances knowledge of the long-term health effects of $PM_{2.5}$ by providing more precise estimates and clearer causal interpretations. These advancements – in exposure assessment, analytical approach, and bias reduction – strengthen confidence in the evidence linking chronic $PM_{2.5}$ ex-

posure to adverse health outcomes, and equip future research with better tools to continue expanding this knowledge base.

Chapter 8

Overview of the research contributions

The five publications presented in the following chapters consist of the core work of my thesis. I led on and am the first author for all of them. The first two full-length articles and a research letter have already been published in peer-reviewed epidemiological and cardiovascular research journals. Another manuscript is currently under revision, while the last research work still needs to be submitted, respectively. The order set in the thesis for the publications reflects both their chronological sequence and the development of my research work. These articles can be read separately as independent original research contributions. However, they also form a coherent body of work.

In this chapter, I provide a summary of the publications. First, I describe the preliminary exposure linkage process necessary to establish epidemiological analyses on the UK Biobank. Second, I present two epidemiological analyses

investigating the association between $PM_{2.5}$ on mortality and hospital admissions. Third, I present an investigation of potential confounding mechanisms affecting UK Biobank epidemiological analyses of long-term environmental exposures. The research letter, discussing a methodological issue related to the research area, is included as last.

8.1 Research Article 1: exposure linkage of pollution to cohort data

The first work, presented in 9, has been published in the Journal of Exposure Science & Environmental Epidemiology[111] and consists of an in-depth and critical illustration of the linkage of outdoor environmental exposure with cohort data, using the UK Biobank dataset as a case-study example. As the first author, I structured the methodological and conceptual phase of the linkage process with all the co-authors, particularly with my supervisor, Prof. Antonio Gasparrini.

This paper introduces an approach to enhancing epidemiological studies of environmental health risks by reconstructing individual-level exposure histories with high precision.

Traditional air pollution models have faced limitations, notably the use of coarse, temporally aggregated exposure metrics or simplified assignment strategies based on sparse monitoring data, which have led to potential exposure misclassification and limited ability to accurately evaluate individual health risks. To address these limitations, the paper introduces and describes a state-of-the-art linkage framework that combines detailed residential histo-

ries of individuals with high-resolution, temporally refined environmental exposure maps (see appendix). The key advancement here is the reconstruction of continuous, individual-level exposure profiles that precisely reflect changes in residential location over time, overcoming the simplistic approaches of earlier studies which typically relied on static, spatially aggregated exposure averages or limited temporal resolutions. By using spatial interpolation methods (such as bilinear interpolation), the framework ensures enhanced accuracy and privacy protection, avoiding the pitfalls of simpler exposure assignment techniques. Moreover, the approach facilitates flexible and precise definitions of exposure windows, tailored to specific epidemiological research designs and questions, enabling more robust analyses of both short-term and long-term health risks. Finally, the resulting linkage framework is not only tailored to UK Biobank – it's generalizable. The procedure can be applied to any cohort with residential histories, providing a template for future researchers to integrate environmental data into health datasets. Importantly, the exposure dataset produced (covering the UK) will be shared with the broader research community.

This paper fulfills Objective 1 of this PhD: *1) to perform a linkage between residential history records and high-resolution $PM_{2.5}$ maps to reconstruct individual-level daily exposure profiles for all the UKB participants.*
In the publication, the written description is enriched with an example, including pseudo-data and informative visualisations, which I developed personally with the assistance of Malcolm Mistry and Antonio Gasparrini. Even though the exposure linkage is focused on highly resolved spatiotemporal exposure data, it can be used for any geolocated exposure dataset. I took the lead in writing the manuscript and acted as the corresponding author during

the peer-review process, addressing the reviewer's comments and making the necessary changes.

8.2 Research Article 2: analysis on long-term association of $PM_{2.5}$ with mortality

The second research paper, originally submitted and recently accepted in Epidemiology is included in 10. It is the first of two publications using the linked exposure data to conduct substantive epidemiological analyses.

This paper provides a innovative analysis of the long-term health effects of $PM_{2.5}$ exposure, particularly focusing on premature mortality for different causes, using data from the UK Biobank. A significant scientific advancement of this research lies in the availability of detailed individual-level daily exposure profiles of $PM_{2.5}$ obtained as an output of objective 1. This refined approach may significantly reduce exposure misclassification compared to earlier studies that relied on coarse exposure assessments or static residential locations. As a preliminary part of the work, I conducted a comparison between the full-cohort Cox analysis and a more computationally efficient nested case-control design (presented in the thesis appendix) which demonstrated that similar results could be obtained with the alternative design. This exercise of evaluating design efficiency vs. validity is a novel contribution, as it provides evidence on whether large biobank analyses can be reliably approximated by sub-sampling methods. Moreover, the paper evaluates and applies state-of-the-art statistical methodologies, including distributed lag

models, to precisely investigate lagged associations, and various windows of exposure. This represents an improvement over prior research methodologies, offering enhanced insight into the temporal dynamics of environmental health impacts. Finally, these analyses are characterized by extensive adjustment for confounding factors at both the individual and contextual levels, thus providing a more realistic understanding of the causal relationships involved.

By being one of the first European studies to incorporate highly resolved, time-varying exposures in a half-million cohort, this analysis directly advances the evidence base beyond the predominantly North American studies of the past. It also provides an updated benchmark of $PM_{2.5}$ mortality risk for the UK, obtained through more sophisticated exposure assignment and modeling than past cohort studies.

This paper partially fulfils Objectives 2 and 3 of this PhD: *2) To implement state-of-the-art designs and statistical methods to investigate health risks associated with $PM_{2.5}$, accounting for confounding mechanisms and assessing complex temporal relationships. 3) To analyse the impact of long-term exposure on premature mortality and hospitalisations for cardio-respiratory outcomes using the large and rich database of the UK Biobank cohort.* As the paper's first author, I conceptualised and coordinated the work with all the co-authors, discussing the study design, research question, and relevant epidemiological and public health issues. I carried out the analysis and developed the corresponding code under the supervision of Prof Antonio Gasparrini. I took the lead in writing the manuscript, assisted by Prof Gasparrini, who acted as the corresponding author during the submission process, addressing and replying to the reviewers' comments.

8.3 Research Article 3: analysis of long-term risks of $PM_{2.5}$ with cardiovascular hospital admissions

The third research paper, published in Environment International, is the second of two publications using the linked exposure data to conduct substantive epidemiological analyses. Using a similar study design and analytical framework developed for Research Article 2, I explored the association between long-term $PM_{2.5}$ exposure and clinical event outcomes. Specifically, I focused on hospital admissions due to various cardiovascular events. This is the first comprehensive long-term investigation on cardiovascular diseases events in the UK. The study it is a unique case in the UK as it leverages highly resolved spatio-temporal $PM_{2.5}$ maps and thoroughly applies state-of-the-art epidemiological methodologies, as described in the previous paragraph. Utilizing the extensive and detailed UK Biobank cohort database, the study offers one of the most comprehensive assessments of multiple cardiovascular outcomes in the UK, including Major Adverse Cardiovascular Events (MACE), myocardial infarction, heart failure, atrial fibrillation, cardiac arrest, and stroke. Importantly, in the analysis I went a step further differentiating subtypes of events – it is the first long-term study to separately examine myocardial infarction and stroke subtypes. By incorporating detailed individual-level confounders (e.g., smoking, waist-to-hip ratio, socioeconomic status, lifestyle factors), alongside area-level covariates (e.g., urban-rural classification, deprivation indices, greenness), the paper sets a new benchmark in addressing potential confounding mechanisms in the air

pollution-cardiovascular diseases relationship.

As the paper's first author, I conceptualised and coordinated the work with all the co-authors, discussing the study design, research question, and relevant epidemiological and public health issues. I carried out the analysis and developed the corresponding code under the supervision of Prof Antonio Gasparrini. I took the lead in writing the manuscript and acted as the corresponding author during the review process, addressing and replying to the reviewer's comments.

8.4 Research Article 4: assessment of confounding mechanisms and adjustment strategies in air pollution epidemiology

The fourth research Article is submitted to the International Journal of Epidemiology. As a further novel aspect of this PhD I present a comprehensive empirical investigation of confounding in cohort analyses for chronic air pollution exposure studies, paired with strategies to address it. Rather than treating confounder adjustment as a black-box step, the research explicitly elucidated how different confounding mechanisms operate in the $PM_{2.5}$–health relationship. By using DAGs alongside empirical analyses, the study mapped out the causal pathways and identified points where spurious associations could arise. This theoretical framework, combined with extensive descriptive correlations in the data, allowed me to anticipate which factors (spatial, temporal, socio-economic, behavioral) needed control. Such a in-depth con-

founding assessment is an innovation in itself, as conventional studies often adjust for available covariates without clearly distinguishing which type of confounding each addresses. Here, by contrast, the roles of spatial context, time trends, and individual lifestyles were separately examined and confirmed in both the conceptual model and the data, yielding clearer guidance on confounder inclusion.

This paper fulfils objective 4 of this PhD: *4) to evaluate confounding mechanisms affecting cohort analyses on long-term exposure to air pollution from theoretical and practical perspectives.* As the paper's first author, I conceptualised and coordinated the work with all the co-authors, discussing the research question and the relevant epidemiological and public health issues. I carried out the analysis and the corresponding code mainly supervised by Prof. Antonio Gasparrini. I took the lead in writing the manuscript and will act as a corresponding author during the review process, addressing and replying to the reviewer's comments.

Chapter 9

Research article 1

Reconstructing individual-level exposures in cohort analyses of environmental risks: an example with the UK Biobank

London School of Hygiene & Tropical Medicine
Keppel Street, London WC1E 7HT
T: +44 (0)20 7299 4646
F: +44 (0)20 7299 4656
www.lshtm.ac.uk

国立大学法人
長崎大学
NAGASAKI UNIVERSITY

RESEARCH PAPER COVER SHEET

Please note that a cover sheet must be completed for each research paper included within a thesis.

SECTION A – Student Details

Nagasaki Student No	59720005	**Title**	Reconstructing individual-level exposures in cohort analyses of environmental risks: an example with the UK Biobank.
LSHTM Student ID No	2004062		
First Name(s)	Jacopo		
Surname/Family Name	Vanoli		
Thesis Title	Assessment of the effect of air pollution on the UK Biobank cohort		
Nagasaki Supervisor(s)	Lina Madaniyazi		
LSHTM Supervisor(s)	Antonio Gasparrini		

If the Research Paper has previously been published please complete Section B, if not please move to Section C.

SECTION B – Paper already published

Where was the work published?	Journal of Exposure Science and Environmental Epidemiology		
When was the work published?	2024		
If the work was published prior to registration for your research degree, give a brief rationale for its inclusion			
Have you retained the copyright for the work?*	**Yes**	Was the work subject to academic peer review?	**Yes**

*If yes, please attach evidence of retention. If no, or if the work is being included in its published format, please attach evidence of permission from the copyright holder (publisher or other author) to include this work.

SECTION C – Prepared for publication, but not yet published

Where is the work intended to be published?	

Please list the paper's authors in the intended authorship order:	
Stage of publication	Choose an item.

SECTION D – Multi-authored work

For multi-authored work, give full details of your role in the research included in the paper and in the preparation of the paper. (Attach a further sheet if necessary)	**I took the lead in carrying out the analysis, structuring and writing up the paper**

SECTION E – Names and affiliations of co-author(s)

Please list all the co-authors' names and their affiliations.

Jacopo Vanoli[1,2], Malcolm N. Mistry[2,3], Arturo De La Cruz[2], Pierre Masselot[2,4], Rochelle Schneider[2,5], Chris Fook Sheng Ng[6], Lina Madaniyazi[1], Antonio Gasparrini[2]

[1]School of Tropical Medicine and Global Health, Nagasaki University, [2]Environment & Health Modelling (EHM) Lab, Department of Public Health Environments and Society, London School of Hygiene & Tropical Medicine, London, UK, [3]Department of Economics, Ca' Foscari University of Venice, Venice, Italy, [4]Centre for Statistical Methodology, London School of Hygiene & Tropical Medicine, London UK, [5]Φ-lab, European Space Agency, Frascati, Italy, [6]Department of Global Health Policy, Graduate School of Medicine, The University of Tokyo

SECTION F

I confirm that all co-authors have agreed that the above paper will be included in my PhD thesis.

Student Signature	[redacted]
Date	9/5/2024

LSHTM Supervisor Signature	[redacted]
Date	05/09/2024

Nagasaki University Supervisor Signature	[redacted]
Date	05/09/2024

Please submit this form to: tmghadmin@ml.nagasaki-u.ac.jp

Journal of Exposure Science & Environmental Epidemiology www.nature.com/jes

PERSPECTIVE OPEN

Check for updates

Reconstructing individual-level exposures in cohort analyses of environmental risks: an example with the UK Biobank

Jacopo Vanoli[1,2]✉, Malcolm N. Mistry[2,3], Arturo De La Cruz Libardi[2], Pierre Masselot[2], Rochelle Schneider[2,4], Chris Fook Sheng Ng[5], Lina Madaniyazi[1] and Antonio Gasparrini[2]

Recent developments in linkage procedures and exposure modelling offer great prospects for cohort analyses on the health risks of environmental factors. However, assigning individual-level exposures to large population-based cohorts poses methodological and practical problems. In this contribution, we illustrate a linkage framework to reconstruct environmental exposures for individual-level epidemiological analyses, discussing methodological and practical issues such as residential mobility and privacy concerns. The framework outlined here requires the availability of individual residential histories with related time periods, as well as high-resolution spatio-temporal maps of environmental exposures. The linkage process is carried out in three steps: (1) spatial alignment of the exposure maps and residential locations to extract address-specific exposure series; (2) reconstruction of individual-level exposure histories accounting for residential changes during the follow-up; (3) flexible definition of exposure summaries consistent with alternative research questions and epidemiological designs. The procedure is exemplified by the linkage and processing of daily averages of air pollution for the UK Biobank cohort using gridded spatio-temporal maps across Great Britain. This results in the extraction of exposure summaries suitable for epidemiological analyses of both short and long-term risk associations and, in general, for the investigation of temporal dependencies. The linkage framework presented here is generally applicable to multiple environmental stressors and can be extended beyond the reconstruction of residential exposures.

IMPACT: This contribution describes a linkage framework to assign individual-level environmental exposures to population-based cohorts using high-resolution spatio-temporal exposure. The framework can be used to address current limitations of exposure assessment for the analysis of health risks associated with environmental stressors. The linkage of detailed exposure information at the individual level offers the opportunity to define flexible exposure summaries tailored to specific study designs and research questions. The application of the framework is exemplified by the linkage of fine particulate matter ($PM_{2.5}$) exposures to the UK Biobank cohort.

Keywords: Epidemiology; Exposure Modeling; Air pollution; Exposure linkage

Journal of Exposure Science & Environmental Epidemiology; https://doi.org/10.1038/s41370-023-00635-w

INTRODUCTION

The role of environmental factors as determinants of health has gained importance in the last decades. Early epidemiological studies have investigated the health impacts of environmental stressors, in particular assessing the mortality risks associated with exposure to air pollutants such as particulate matter [1]. The evidence has been subsequently strengthened and extended to a variety of other exposures and outcomes [2, 3]. Emergent research also suggests health risks associated with other environmental exposures, such as other pollutants such as nitrogen oxides, temperature, pollen, and other chemicals [2, 4], as well as for a variety of health outcomes, including communicable and non-communicable disease [5].

A known problem in this research area is that most environmental stressors, while affecting entire populations and generating considerable health burdens, are usually associated with relatively low health risks at the individual level. Estimating such associations therefore requires large epidemiological studies. With few exceptions [6], early investigations relied on administrative databases with limited individual information and were often based on ecological designs [7]. Nowadays, new opportunities are offered by the availability of large population-based cohorts that match the recruitment of a high number of participants with the detailed reconstruction of individual information through linkage across multiple databases. Recent endeavours, such as the European EPIC study, the UK Biobank [8], and the Japanese JECS include the collection of detailed questionnaires and physical measurements, through which it is possible to explore small variations in susceptibility due to lifestyles, genetic traits, and other individual and contextual characteristics.

[1]School of Tropical Medicine and Global Health, Nagasaki University, Nagasaki, Japan. [2]Environment & Health Modelling (EHM) Lab, Department of Public Health Environments and Society, London School of Hygiene & Tropical Medicine, London, UK. [3]Department of Economics, Ca' Foscari University of Venice, Venice, Italy. [4]Φ-lab, European Space Agency, Frascati, Italy. [5]Department of Global Health Policy, Graduate School of Medicine, The University of Tokyo, Tokyo, Japan. ✉email: jacopo.vanoli@lshtm.ac.uk

Received: 25 April 2023 Revised: 8 December 2023 Accepted: 12 December 2023
Published online: 08 January 2024

A related problem is represented by the exposure assessment. Direct personal monitoring of environmental exposures is unfeasible for large-scale studies across long periods of time, and therefore outdoor levels at residential locations are typically used as a proxy for personal exposure. Early cohort studies made use of data from sparse monitoring stations, which can result in misclassification and reduced exposure contrasts [1, 9], more so for exposure that features high spatial and/or temporal variability such as air pollution. Nowadays, exposure modelling techniques offer valuable solutions with improved prediction accuracy and coverage. For instance, modern methodologies can combine multi-domain predictors in sophisticated analytical models to derive high-resolution spatio-temporal maps over large regions [10]. These methods have been previously used to harmonise the exposure assignment to large population-based cohorts in North America [11] and Europe [12].

Such models nonetheless do not always produce temporally disaggregated measures [13], required for assessing short-term risks. Other studies have assigned annual exposure averages, but without accounting for residential changes and potential long-lagged associations with past exposures [6]. More informative and accurate exposure summaries can be defined by reconstructing the complete exposure history for each cohort participant. This extension offers the possibility to examine other aspects such as multiple association timescales and windows of susceptibility. However, this extension presents important methodological, logistical, and practical issues.

In this contribution, we present a currently applied framework for the linkage of highly resolved outdoor environmental exposures to large cohorts using individual residential information. The illustration provides the opportunity to discuss methodological aspects and technical requirements, as well as specific problems such as privacy constraints. We exemplify this process by assigning exposures to air pollution to the UK Biobank cohort, a large prospective study involving more than half a million participants. The article outlines a number of steps needed to generate individual-level exposure profiles, and finally to derive exposure summaries consistent with alternative study designs and research questions.

MATERIALS AND METHODS

UK Biobank

The UK Biobank cohort is a longitudinal study that has involved adults aged 40–69 at recruitment in the United Kingdom between 2006 and 2010 [8]. Overall, 503,325 participants were recruited and each of them attended an assessment centre and completed questionnaires on their socio-economic aspects, lifestyle factors, and medical history, among other information. They also underwent a wide range of physical measures, as well as the collection of biological samples. The study is periodically enriched with follow-up assessments, new sources of data originating from research projects, and updates from external databases. These comprise the linkage with electronic health records (EHR) and national health system registers, including death and cancer occurrences, hospitalisations and primary care visits. Information on environmental exposures currently available in the UK Biobank is represented by annual averages of air pollutants and noise for single years between 2006 and 2010. Air pollution measures are limited to a sub-group of participants and obtained from Europe-wide land-use regression models [14].

The linkage of new environmental data to cohort participants necessitates three sources of information, exemplified by the pseudo-data illustrated in Table 1. These simulated data are used in this and the next sections to describe the linkage process and epidemiological analyses. The first piece of information is about the baseline cohort information, illustrated in Table 1a. These data are represented here by the enrolment and last follow-up dates for each participant, identified by a pseudo-code. This usually is linked to other information collected at the baseline or during follow-up assessments, such as personal characteristics and socio-economic factors, which are not shown here. The

Table 1. Example of pseudo cohort data, including **a** baseline cohort information, **b** health outcomes, and **c** residential histories.

(a) Cohort info

Subject ID	Enrolment date	Last follow-up date
1	May 1, 2007	March 12, 2017
2	April, 14, 2009	September 25, 2019
3	November 23, 2006	Present

(b) Inpatient visit outcomes table by subject

Subject ID	ICD	Date
1	E11	April 23, 2012
1	I20	July 4, 2013
1	I21	September 30, 2016
2	C34	February 24, 2010
3	J40	March 14, 2007
3	J41	April 11, 2008
3	J43	May 22, 2009

(c) Residential histories

Subject ID	Location ID	Start date	End date	Easting	Northing
1	Loc_12	April 1, 2005	May 22, 2012	515,200	184,800
1	Loc_43	May 23, 2012	March 12, 2017	384,800	394,100
2	Loc_92	December 18, 2007	September 3, 2009	342,700	387,100
2	Loc_6	September 4, 2009	April 3, 2017	528,100	105,600
2	Loc_24	April 4, 2017	September 25, 2019	459,900	450,700
3	Loc_87	November 20, 1994	Present	177,500	314,500

second piece of information concerns the health data, some of which is accessible to UK Biobank researchers through a standard application. For instance, the main database includes inpatient records of the first occurrences of a series of clinical adverse events. An example with pseudo-data is provided in Table 1b, including the same pseudo-IDs of the subject, as well as the ICD-10 codes and dates of the events.

The final piece of information is the residential histories of the subjects. In the UK Biobank, these are limited-access data, represented by the dates and locations of the participants' residential addresses, where the location represents the centroid of a 1 km and 100 m buffer that contains the exact location. These data were collected during the baseline interview and are ongoingly updated via self-report or new registration to general practices of the National Health Service (NHS). Residential pseudo-data are shown in Table 1c, including pseudo-IDs for subjects and locations, and start/end dates of the period the subject stayed at each address, alongside the corresponding geographical coordinates (in Northing-Eastings values of the British National Grid).

Spatio-temporal exposure maps

Advances in exposure assessment have been achieved through important developments in two areas. First, the increasing availability of data resources with high spatial and temporal resolution and extended coverage, in particular from remote sensing sources. Second, the provision of innovative analytical techniques, for instance, machine learning algorithms or atmospheric and climate models with increasingly better performance and reliability. These technological advancements make it possible to produce fine-scale spatio-temporal maps of environmental exposures applicable in population-based epidemiological studies [15]. These state-of-the-art tools have rapidly substituted classical exposure assessment methods, such as the assignment to the closest monitoring station or traditional land-use regression models, as the latter fail to provide accurate estimates for large areas and over long periods of time [16].

In this contribution, we consider a dataset that is currently used to assign daily exposures to fine particulate matter ($PM_{2.5}$, in μg/m^3) to the participants locations of the UK Biobank. This product was generated by a multi-stage machine learning model that was applied to predict daily $PM_{2.5}$ concentrations in a 1×1 km grid across Great Britain during the period 2008–2018. The model was trained using data from 581 monitoring stations, using a long list of spatial and spatio-temporal predictors including remote sensing satellite observations, traffic data, weather simulations, road characteristics, and land-use information, among others. The model had a good overall performance, with a cross-validated R^2 of 0.767. Details are provided elsewhere [16].

This resource is used in the next sections to exemplify the linkage process of $PM_{2.5}$ measures to participants of the UK Biobank.

Spatial linkage (Step 1)

Geographical information systems (GIS) have become a staple technique for constructing environmental databases. In this context, GIS provide a binding framework between environmental measures and cohort data collected at the individual level, combining different layers of information to a single point in space [17]. These techniques are employed in epidemiological analyses by overlying geographical reference grids over which the investigators can jointly map exposure information with individual or area-level variables. This allows maximising the available information by downscaling or upscaling measurements across levels of aggregation, as well as combining measurements across space and time.

We discuss the application of GIS techniques and related problems by illustrating the linkage of environmental exposures to the UK Biobank. The cohort database includes the locations of the residential addresses of each participant. An example is provided in Fig. 1, which shows the $PM_{2.5}$ levels for one day from the 1×1 km gridded spatio-temporal map presented in the previous section. The map also includes the three residential addresses for Subject 1 listed in Table 1b, and for one address, it adds a magnified detail of the 1×1 km cells surrounding the location.

A simple linkage option is to assign the value of the grid cell containing the location. However, this option has two main drawbacks. First, it does not account for the information of the neighbouring cells, which can complement the cell-level measurement with details on the small-scale variability and improve the exposure assignment. Second, and more importantly, the direct linkage of cell-specific values can result in potential privacy breaches described above by allowing back-tracing of the location using geographic information from the original gridded environmental data, if this is publicly available and at sufficiently high resolution.

In lieu of the simple linkage approach described above, other methods of varying complexity can be used and the choice depends on the type of exposure data and the underlying

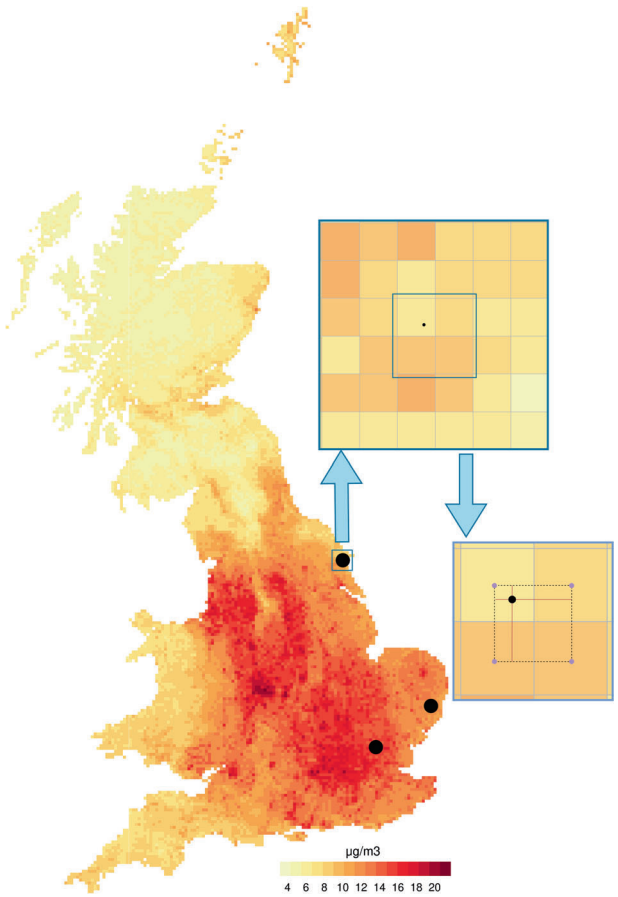


Fig. 1 The maps display $PM_{2.5}$ levels on a specific day over Great Britain, with three locations (large black dots) that represent the residential addresses of a specific subject (ID 2 in Table 1). The magnified area on top represents the exact location at higher resolution, surrounded by the four nearest centroids (small indigo dots) of the overlaid $PM_{2.5}$ grid. Without interpolation, the residential exposure value (small black dot) would be represented by the value of the nearest centroid. The magnified area below illustrates the process of reconstructing the residential value as a bilinear interpolation of the four nearest centroids.

objective of data linkage. For example, in the presence of ground monitor data, a simple strategy would be to assign exposure as the inverse-distance weighted average of the nearby monitors. For gridded exposure data, established routines such as simple spatial averaging, bilinear and kriging interpolation exist in the two-dimensional case, while more specific methods have been investigated more recently as a consequence of the raise of new forms of spatial data [18]. Here, we propose the use of the bilinear interpolation, which consists of a repeated linear interpolation across the two geographical dimensions and it is graphically represented in Fig. 1. We deem this method to be an effective but simple option, among the others, for several reasons. The process addresses the two drawbacks of the simpler linkage described above: first, it preserves the exposure information by spatially combining measurements across multiple grid cells. Second, and more importantly, it generates a continuous exposure field with values that cannot be linked back to the original sources, preventing the identification of the residential locations even when using highly resolved and public exposure databases. Compared to other interpolation methods, bilinear interpolation does not require a choice of the parameters (e.g., search radius or number of neighbours) and it is more accurate than simple spatial averaging as it accounts for the distances among the points in the computation of the interpolated value [19]. Moreover, its deterministic nature makes it computationally inexpensive even for very large datasets, for instance in comparison to kriging [20]. Finally, bilinear interpolation is commonly implemented in data analysis and geographical software and therefore easy to apply. It must be highlighted that, regardless of the method, the accuracy of this linkage would depend on the spatial resolution of the original exposure data, and the precision of the coordinates for the locations.

Reconstruction of individual-level exposure series (Step 2)

The linkage-interpolation operation in the previous section can be performed for each residential location of each participant of the cohort. The output data, combined with the residential histories, allow reconstructing subject-specific series representing individual exposure profiles.

This step is illustrated in Fig. 2 for Subject 2 in our case study. Specifically, the residential histories of this subject reported in Table 1c, combined with the interpolated series for the three residential locations obtained following the procedure in Fig. 1, allow extracting blocks of exposure series corresponding to the timeline of each subject's residence at specific addresses. These blocks are then merged into a single individual series that represents a detailed residential exposure profile for an individual, accounting for exposure levels experienced at different locations during a defined time interval.

Definition of individual summaries for epidemiological studies (Step 3)

The reconstruction of exposure profiles in the previous section offers detailed individual-level time series characterised by a fine temporal disaggregation, allowing the definition of various exposure summaries. In epidemiological analyses, this is of particular relevance as such summaries can be flexibly tailored to the specific research questions and study designs, resulting in more informative inferential procedures and reducing exposure misclassification.

The definition of the exposure summaries first requires assumptions on the temporal dependency between exposure and outcomes, determined by underlying biological mechanisms. Two intertwined aspects are particularly relevant: the timescale of the association and the related exposure window. The former differentiates short-term risks associated with daily variation from long-term effects due to chronic exposures experienced over years or decades. The latter determines the maximal temporal interval over which the exposure exerts its action, within a specific timescale.

We use our case study to illustrate the definition of exposure summaries for two different study designs for individual-level data: a survival analysis based on Cox proportional hazard models to assess long-term effects [21], and a case-crossover analysis to investigate short-term associations [22]. The two examples are represented in Fig. 3, using the pseudo-data related to specific health events in Table 1b.

The Cox proportional hazard model is based on a between-subject comparison, defining separate risk sets for each event. Each risk set includes the case subject as well as a series of control subjects who are at risk at the time of the event. An example of a single risk set is shown at the top of Fig. 3. The composition of the risk set depends on the time axis of interest, which in this case is represented by the age of the subjects. The controls are therefore sampled when they reach the same age that the case had when experiencing the event. For each subject, we retrieve their exposure history backwards with a lag period equal to the exposure window, and therefore define the related exposure summary.

A case-crossover design follows a similar extraction procedure. However, in contrast to the survival model above, the latter is based on a within-subject comparison, and the case and controls are represented by different times within the follow-up period of the same subject. Several control sampling schemes have been proposed in the literature [23] with the most common being the time-stratified scheme with controls sampled within pre-specified strata. An example with three subjects representing three separate risk sets with an exposure window of four days (lag 0–3) is provided at the bottom of Fig. 3.

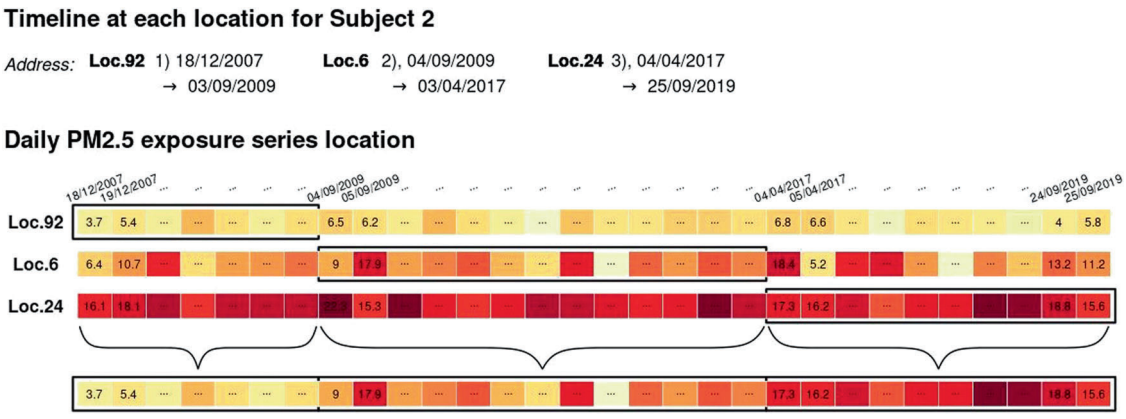


Fig. 2 The top three series represent the sequences of daily exposures at the residential addresses of subject ID 2. At the bottom, the final subject-specific exposure series is assembled by concatenating the three series above based on the respective residential periods.

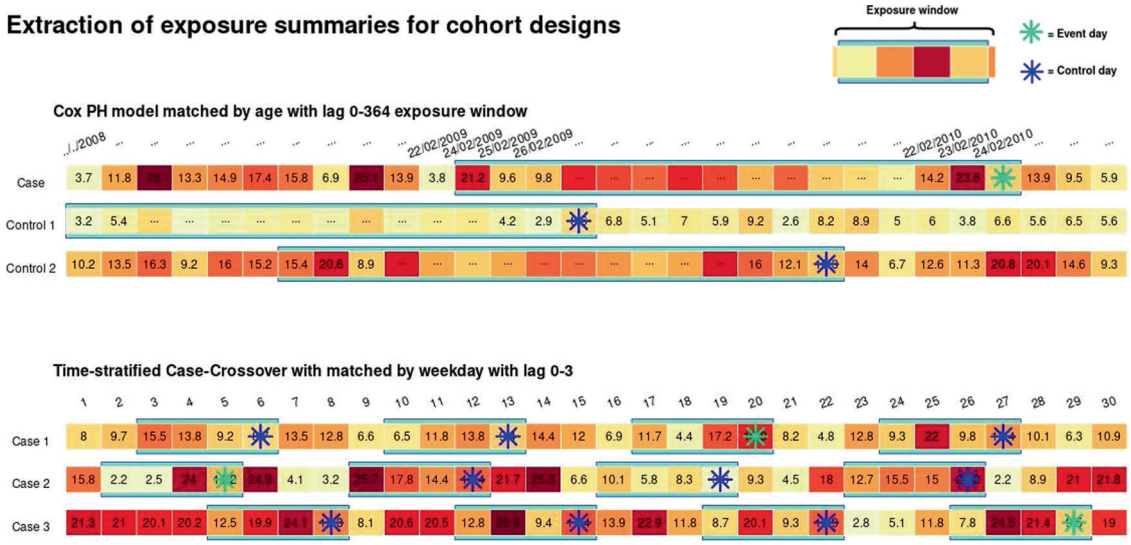


Fig. 3 The graph presents the use of the exposure data in two examples of study designs used in environmental epidemiology. The top figure illustrates a risk set within a study on the incidence of lung cancer (ICD-10: C34) with a case (subject 2) and controls matched by age used in a Cox proportional hazard model to estimate long-term risks. The event (aquamarine star) and control (blue star) times are used to reconstruct backwards the exposure profiles in the three subjects, defined as 365-day (lag 0–364) averages of $PM_{2.5}$ (light blue boxes). The bottom figure displays the same process to define risk sets for a time-stratified case-crossover to estimate short-term risks. The graph shows three separate subjects (unrelated to Table 1) with the event (aquamarine star) and controls (blue star) days matched on the day of the week in the same month, with exposure profiles defined as averages of lag 0–3.

The availability of finely stratified temporal profiles allows higher precision in the definition of the exposure windows, before any potential aggregations are performed. For instance, multiple lag terms can be defined using daily, monthly, or yearly strata, thus allowing the application of distributed lag models over different timescales [24].

DISCUSSION

This article describes a framework to process and link environmental exposures to cohort studies. The methodology can be applied to retrieve detailed individual-level exposure profiles, hence allowing the application of flexible epidemiological study designs to investigate health risks associated with environmental stressors. The paper conceptualises several steps and methodological aspects, with illustration in a case study featuring the UK Biobank cohort using simplified pseudo-datasets. The framework has broad applications and can be used to complement cohort databases with high-resolution spatio-temporal exposure measurements, enabling to investigate complex aetiological questions between environmental factors and health.

This work can contribute to clarify and improve on current limitations in the research field. An example is offered by recent cohort analyses of associations between low levels of air pollution with mortality and morbidity conducted in the USA, Canada, and Europe [6]. These investigations applied state-of-the-art methodologies to large population-based cohort databases, representing milestones in air pollution epidemiology. Specifically, the North American studies examined health risks associated with several air pollutants by reconstructing exposures with resolved spatial predictions and various temporal disaggregation [11]. However, these cohort analyses often relied on administratively collected cohort data whereby, due to privacy constraints, exposure information could only be matched to large administrative areas. In contrast, recent multi-cohort European studies [13] took advantage of exposure models with high spatial resolution and linkage at residential level. However, the exposure data was not temporally disaggregated, and the analyses relied on simple exposure summaries based on averages for specific numbers of years, preventing the investigation of complex temporal dependencies. The framework presented here, given the availability of the data, helps addressing these limitations, providing a privacy-protecting approach to safely link resolved spatio-temporal exposure maps to large databases with rich individual information, thereby improving the design of cohort studies.

The example based on the UK Biobank cohort also highlights some practical problems. First, our choice of the interpolating method was based on practical criteria, but in general this decision would benefit from rigorous comparisons, for instance based on statistical goodness of fit measures [19]. Second, the linkage procedure exemplified necessitates information on residential mobility. Currently, in the UK Biobank such data is only reconstructed from participants' self-reports and NHS contacts. This process is error-prone and can entail exposure misclassification. Third, the accuracy of the exposure assessment depends on the quality and resolution of the spatio-temporal exposure models. In our example, we demonstrated a linkage with gridded databases of pollution derived from moderate-to-high predictive performance, which similarly provides an imperfect characterisation of exposure levels. Finally, even when accurately representing residential levels, outdoor estimates are only a proxy of the actual personal exposures.

Nonetheless, the framework described here offers a template for future developments to address current limitations and overcome new challenges. Most importantly the approach can be extended beyond the linkage of residential measurements, for instance incorporating activity-based models or personal monitoring campaigns to improve individual exposure assessment in different environments [25]. This is relevant as hyperlocal exposure models are increasingly deployed in urban settings with the aim of addressing environmental disparities [26] and the environmental datasets can be made publicly available to researchers [27]. Finally, the assignment of individual-level exposure profiles can be replicated for multiple stressors. This will allow the investigation of health risks associated with the bulk of environmental exposures, consistent with the notion and

research paradigm of the exposome [28]. In this context, the linkage framework we illustrated can be applied and further developed to finely reconstruct detailed exposure information across large cohorts and long study periods, while at the same time preventing confidentiality breaches by providing bespoke exposure levels that cannot be traced back to the original data.

DATA AVAILABILITY

The code and example data for replicating the illustrative example are made available upon request from the corresponding author. The analysis was performed in the R software environment.

REFERENCES

1. Pope CA III, Thun MJ, Namboodiri MM, Dockery DW, Evans JS, Speizer FE, et al. Particulate air pollution as a predictor of mortality in a prospective study of U.S. adults. Am J Respir Crit Care Med. 1995;151:669–74.
2. Atkinson RW, Butland BK, Anderson HR, Maynard RL. Long-term concentrations of nitrogen dioxide and mortality: a meta-analysis of cohort studies. Epidemiology. 2018;29:460–72.
3. Jie C, Sophia R, Kees de H, Maciej S, Zorana JA, Richard A, et al. Long-term exposure to fine particle elemental components and natural and cause-specific mortality: a pooled analysis of eight European cohorts within the ELAPSE Project. Environ Health Perspect. 2021;129:047009.
4. Kitinoja MA, Hugg TT, Siddika N, Rodriguez Yanez D, Jaakkola MS, Jaakkola JJK. Short-term exposure to pollen and the risk of allergic and asthmatic manifestations: a systematic review and meta-analysis. BMJ Open. 2020;10:e029069.
5. World Health Organization. WHO global air quality guidelines. Particulate matter (PM 2.5 and PM 10), ozone, nitrogen dioxide, sulfur dioxide and carbon monoxide. Geneva: WHO; 2021.
6. Pope CA III, Coleman N, Pond ZA, Burnett RT. Fine particulate air pollution and human mortality: 25+ years of cohort studies. Env Res. 2020;183:108924.
7. Samet JM, Dominici F, Curriero FC, Coursac I, Zeger SL. Fine particulate air pollution and mortality in 20 U.S. cities, 1987-94. N Engl J Med. 2000;343:1742–9.
8. Sudlow C, Gallacher J, Allen N, Beral V, Burton P, Danesh J, et al. UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. PLoS Med. 2015;12:e1001779.
9. Dockery DW, Pope CA, Xu X, Spengler JD, Ware JH, Fay ME, et al. An association between air pollution and mortality in six U.S. cities. N Engl J Med. 1993;329:1753–9.
10. Chen J, de Hoogh K, Gulliver J, Hoffmann B, Hertel O, Ketzel M, et al. Development of Europe-wide models for particle elemental composition using supervised linear regression and random forest. Env Sci Technol. 2020;54:15698–709.
11. Di Q, Wang Y, Zanobetti A, Wang Y, Koutrakis P, Choirat C, et al. Air pollution and mortality in the Medicare population. N Engl J Med. 2017;376:2513–22.
12. Eeftens M, Beelen R, de Hoogh K, Bellander T, Cesaroni G, Cirach M, et al. Development of land use regression models for PM2.5, PM2.5 absorbance, PM10 and PMcoarse in 20 European study areas; results of the ESCAPE project. Environ Sci Technol. 2012;46:11195–205.
13. Strak M, Weinmayr G, Rodopoulou S, Chen J, de Hoogh K, Andersen ZJ, et al. Long term exposure to low level air pollution and mortality in eight European cohorts within the ELAPSE project: pooled analysis. BMJ 2021;374:n1904.
14. Biobank U. UKB environmental exposures link. 2014. https://biobank.ndph.ox.ac.uk/ukb/ukb/docs/EnviroExposEst.pdf.
15. Song W, Wu C. Introduction to advancements of GIS in the new IT era. Ann Gis 2021;27:1–4.
16. Schneider R, Vicedo-Cabrera AM, Sera F, Masselot P, Stafoggia M, de Hoogh K, et al. A satellite-based spatio-temporal machine learning model to reconstruct daily PM2.5 concentrations across Great Britain. Remote Sens. 2020;12:3803.
17. Nuckols JR, Ward MH, Jarup L. Using geographic information systems for exposure assessment in environmental epidemiology studies. Environ Health Perspect. 2004;112:1007–15.
18. Comber A, Zeng W. Spatial interpolation using areal features: a review of methods and opportunities using new forms of data with coded illustrations. Geogr Compass. 2019;13:e12465.
19. Wong DW, Yuan L, Perlin SA. Comparison of spatial interpolation methods for the estimation of air quality data. J Expo Sci Environ Epidemiol. 2004;14:404–15.
20. Randolph BC. Extending kriging methods to large datasets with applications to California groundwater data. UCLA; 2017. https://escholarship.org/uc/item/23d1w7vv. Accessed 23 Oct 2023.
21. Andersen PK, Gill RD. Cox's regression model for counting processes: a large sample study. Ann Stat. 1982;10:1100–20.
22. Maclure M. The case-crossover design: a method for studying transient effects on the risk of acute events. Am J Epidemiol. 1991;133:144–53.
23. Mittleman MA, Mostofsky E. Exchangeability in the case-crossover design. Int J Epidemiol. 2014;43:1645–55.
24. Gasparrini A. Modeling exposure–lag–response associations with distributed lag non-linear models. Stat Med. 2014;33:881–99.
25. Larkin A, Hystad P. Towards personal exposures: how technology is changing air pollution and health research. Curr Env Health Rep. 2017;4:463–71.
26. Zhang Y, Smith SJ, Bell M, Mueller A, Eckelman M, Wylie S, et al. Pollution inequality 50 years after the Clean Air Act: the need for hyperlocal data and action. Environ Res Lett. 2021;16:071001.
27. Wang A, Mora S, Machida Y, deSouza P, Paul S, Oyinlola O, et al. Hyperlocal environmental data with a mobile platform in urban environments. Sci Data. 2023;10:524.
28. Vrijheid M. The exposome: a new paradigm to study the impact of environment on health. Thorax. 2014;69:876–8.

ACKNOWLEDGEMENTS

This research has been conducted using the UK Biobank Resource under application number 56431. This work was part of a PhD project supported by the Nagasaki University WISE Programme.

AUTHOR CONTRIBUTIONS

Study conception and design: AG and JV. Draft manuscript preparation: JV and AG. All authors provided substantial critical input to improve the manuscript, and all authors approved the final draft.

FUNDING

This work was supported by the Medical Research Council-UK (Grant IDs: MR/Y003330/1 and MR/R013349/1), the European Union's Horizon 2020 Project Exhaustion (Grant ID: 820655), Nagasaki University "Doctoral Program for World-leading Innovative and Smart Education" for Global Health (WISE), KENKYU SHIDO KEIHI ("the Research Grant") and KYOIKU KENKYU SHIEN KEIHI ("the Stipend").

COMPETING INTERESTS

The authors declare no competing interests.

ADDITIONAL INFORMATION

Correspondence and requests for materials should be addressed to Jacopo Vanoli.

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

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

Chapter 10

Research article 2

Long-term associations between time-varying exposure to ambient $PM_{2.5}$ and mortality risks: an analysis of the UK Biobank

London School of Hygiene & Tropical Medicine
Keppel Street, London WC1E 7HT
T: +44 (0)20 7299 4646
F: +44 (0)20 7299 4656
www.lshtm.ac.uk

国立大学法人
長崎大学
NAGASAKI UNIVERSITY

RESEARCH PAPER COVER SHEET

Please note that a cover sheet must be completed for each research paper included within a thesis.

SECTION A – Student Details

Nagasaki Student No	59720005	**Title**	Long-term associations between time-varying exposure to ambient PM2.5 and mortality: an analysis of the UK Biobank.
LSHTM Student ID No	2004062		
First Name(s)	Jacopo		
Surname/Family Name	Vanoli		
Thesis Title	Assessment of the effect of air pollution on the UK Biobank cohort		
Nagasaki Supervisor(s)	Lina Madaniyazi		
LSHTM Supervisor(s)	Antonio Gasparrini		

If the Research Paper has previously been published please complete Section B, if not please move to Section C.

SECTION B – Paper already published

Where was the work published?			
When was the work published?			
If the work was published prior to registration for your research degree, give a brief rationale for its inclusion			
Have you retained the copyright for the work?*	Choose an item.	Was the work subject to academic peer review?	Choose an item.

*If yes, please attach evidence of retention. If no, or if the work is being included in its published format, please attach evidence of permission from the copyright holder (publisher or other author) to include this work.

SECTION C – Prepared for publication, but not yet published

Where is the work intended to be published?	Epidemiology

Please list the paper's authors in the intended authorship order:	Jacopo Vanoli[1,2], Arturo de la Cruz[2], Francesco Sera[3] , Massimo Stafoggia[4], Pierre Masselot[2,9], Malcolm Mistry[2,5], Sanjay Rajagopalan[6], Jennifer Quint[7], Chris Fook Sheng Ng[8], Lina Madaniyazi[1], Antonio Gasparrini[2,9]
Stage of publication	**In press**

SECTION D – Multi-authored work

For multi-authored work, give full details of your role in the research included in the paper and in the preparation of the paper. (Attach a further sheet if necessary)	**I took the lead in carrying out the analysis, structuring and writing up the paper**

SECTION E – Names and affiliations of co-author(s)

Please list all the co-authors' names and their affiliations.

Jacopo Vanoli[1,2], Arturo de la Cruz[2], Francesco Sera[3] , Massimo Stafoggia[4], Pierre Masselot[2,9], Malcolm Mistry[2,5], Sanjay Rajagopalan[6], Jennifer Quint[7], Chris Fook Sheng Ng[8], Lina Madaniyazi[1], Antonio Gasparrini[2,9]

[1]School of Tropical Medicine and Global Health, Nagasaki University, Nagasaki, Japan, [2]Department of Public Health Environments and Society, London School of Hygiene and Tropical Medicine, London, UK, [3]Department of Statistics, Computer Science, Applications, University of Florence, Florence, Italy, [4]Department of Epidemiology, Lazio Region Health Service, Rome, Italy, [5]Department of Economics, Ca' Foscari University of Venice, Venice, Italy, [6]Harrington Heart and Vascular Institute, University Hospitals Cleveland Medical center, Cleveland, OH, 44106, USA, [7]National Heart and Lung Institute, Imperial College London, London, UK, [8]Department of Global Health Policy, Graduate School of Medicine, The University of Tokyo, Tokyo, Japan, [9]Centre for Statistical Methodology, London School of Hygiene & Tropical Medicine, London UK

SECTION F

I confirm that all co-authors have agreed that the above paper will be included in my PhD thesis.

Student Signature	[redacted]
Date	9/5/2024

LSHTM Supervisor Signature	[redacted]

Date	05/09/2024

Nagasaki University Supervisor Signature	[redacted]
Date	05/09/2024

Please submit this form to: tmghadmin@ml.nagasaki-u.ac.jp

Long-term associations between time-varying exposure to ambient $PM_{2.5}$ and mortality: an analysis of the UK Biobank

Jacopo Vanoli[1], Arturo de la Cruz[1], Francesco Sera[1,2], Massimo Stafoggia[3], Pierre Masselot[1], Malcolm N. Mistry[1,4], Sanjay Rajagopalan[5,6], Jennifer K Quint[7], Chris Fook Sheng Ng[8,9], Lina Madaniyazi[9], Antonio Gasparrini[1*]

[1] Environment & Health Modelling (EHM) Lab, Department of Public Health Environments and Society, London School of Hygiene & Tropical Medicine, London, United Kingdom

[2] Department of Statistics, Computer Science and Applications "G. Parenti," University of Florence, Florence, Italy

[3] Department of Epidemiology, Lazio Region Health Service, ASL Roma 1, Rome, Italy

[4] Department of Economics, Ca' Foscari University of Venice, Venice, Italy

[5] Harrington Heart and Vascular Institute, University Hospitals Cleveland Medical Center, Cleveland, OH Ohio, United States of America

[6] School of Medicine, Case Western Reserve University, Cleveland, Ohio, United States of America

[7] School of Public Health, Imperial College London, London, United Kingdom

[8] Department of Global Health Policy, Graduate School of Medicine, The University of Tokyo, Tokyo, Japan

[9] School of Tropical Medicine and Global Health, Nagasaki University, Nagasaki, Japan

*Correspondence to: Prof Antonio Gasparrini, London School of Hygiene & Tropical Medicine, 15-17 Tavistock Place WC1H 9SH, London, United Kingdom. Email: antonio.gasparrini@lshtm.ac.uk; Telephone: +44 (0)20 7927 2406; ORCID: https://orcid.org/0000-0002-2271-3568.

Running title: Long-term mortality risks of $PM_{2.5}$ in the UK Biobank

Abstract

Background: Evidence for long-term mortality risks of $PM_{2.5}$ comes mostly from large administrative studies with incomplete individual information and limited exposure definitions. Here we assess $PM_{2.5}$-mortality associations in the UK Biobank cohort using detailed information on confounders and exposure.

Methods: We reconstructed detailed exposure histories for 498,090 subjects by linking residential data with high-resolution $PM_{2.5}$ concentrations from spatio-temporal machine learning models. We split the time-to-event data and assigned yearly exposures over a lag window of eight years. We fitted Cox proportional hazard models with time-varying exposure controlling for both contextual and individual-level factors, as well as trends. In secondary analyses, we inspected the lag structure using distributed lag models (DLMs) and compared results with alternative exposure sources and definitions.

Results: In fully-adjusted models, an increase of 10 μg/m³ in $PM_{2.5}$ was associated with hazard ratios (HRs) of 1.27 (95%CI: 1.06-1.53), 1.24 (1.03-1.50), 2.07 (1.04-4.10) and 1.66 (0.86-3.19) for all-cause, non-accidental, respiratory, and lung cancer mortality, respectively. We found no evidence of associations with cardiovascular deaths (HR=0.88, 95%CI: 0.59-1.31). We identified strong confounding effects by both contextual and individual-level lifestyle factors. The DLM analysis suggested potential differences in relevant exposure windows across mortality causes. Using more informative exposure summaries and sources resulted in higher risk estimates.

Conclusions: We found associations of long-term $PM_{2.5}$ exposure with all-cause, non-accidental, respiratory, and lung cancer mortality, but not with cardiovascular causes. This study benefits from finely-reconstructed time-varying exposures and extensive control for confounding, providing further support to a plausible causal link between long-term $PM_{2.5}$ and mortality .

Keywords: particulate matter; mortality; cohort study; UK Biobank; distributed lag models; machine learning.

Introduction

Chronic exposure to fine particulate matter ($PM_{2.5}$) is recognized as a major risk factor for human health. The epidemiological literature has focused particularly on non-accidental mortality, for which now extensive evidence supports the existence of a causal link.[1-3] However, strong associations have also been found for cause-specific outcomes, including cardiovascular, respiratory, and lung cancer mortality,[4-6] with recent evidence extending the link to other outcomes such as metabolic and kidney diseases as well as neurological disorders.[7-9] These associations were shown to persist at low $PM_{2.5}$ concentrations (below 10-12 $\mu g/m^3$), prompting the World Health Organization (WHO) to revise the recommended annual limit downward to 5 $\mu g/m^3$.

The large number of studies on the long-term impact of $PM_{2.5}$ on mortality has been summarised by various meta-analyses.[10,11] The main evidence, in terms of sample size and population representativeness, comes from large administrative cohorts in North America, including the historical Harvard Six Cities and the American Cancer Society (ACS) cohorts,[12] and more recently the US Medicare,[13] and the Canadian Census Health and Environment Cohorts (CanCHEC),[14] as well as multi-centre studies conducted in Europe, for instance within the ESCAPE and ELAPSE projects.[15-17]

Despite this evidence base, there are limitations. First, most of the large studies are based on administrative cohorts with partial information on individual-level characteristics.[13,14,17] Specifically, such analyses ensured control of socio-economic and demographic attributes such as age, sex, income, and education, but they lacked information on lifestyle and other health-related factors. These characteristics, including smoking, alcohol consumption, and physical exercise, as well as body-to-mass index, are among the strongest risk factors for mortality and have a strong potential for confounding.

Second, a critical feature of these studies of long-term PM2.5 and mortality is the exposure assignment process. Recent cohort studies took advantage of state-of-the-art exposure models that provide fine spatio-temporal resolution and accuracy for retrospective assignment of time-varying exposures.[18,19] However, most analyses relied on simple exposure definitions, for instance using same-year $PM_{2.5}$ levels,[13] assigning time-constant values,[15-17] or using single pre-determined exposure windows.[14] Previous studies have stressed the sensitivity of the estimates to the spatial and temporal scale of the exposure assignment,[20] and a more informative definition of the exposure-risk summaries, accounting for time-varying levels and individual residential histories, can lead to improved health risk estimates and knowledge of aetiological mechanisms.

Third, both meta-analyses and pooled studies indicated a degree of heterogeneity in the estimates of health risks associated with long-term exposure to $PM_{2.5}$.[10,11,17] While part of it can be explained by differences in exposure assessment as well as residual confounding, real variations in risks can also be related to particulate composition or underlying vulnerability of populations. Country and region-specific assessments are therefore needed for reliable health risk estimation at the local level. However, evidence from the UK is scarce, with only a few individual-level studies conducted using administrative databases and population-based cohorts.[21-23]

In this contribution, we assessed long-term risks for all-cause and cause-specific mortality associated with exposure to low levels of $PM_{2.5}$ in the UK Biobank cohort. The study benefits from the analysis of a large cohort database with detailed individual characteristics, the use of detailed exposure histories assigned from high-resolution machine learning models and accounting for residential mobility, and the application of state-of-the-art analytical methods to investigate temporal relationships.

Methods

The UK Biobank

The UK Biobank (UKB) is a British prospective cohort study that has enrolled more than half a million participants aged 40 to 69 years between 2006 and 2010.[24] At enrolment, the participants underwent a first in-person assessment in one of the 22 assessment centres located across Great Britain (England, Scotland, and Wales). The assessment consisted of multiple questionnaires collecting information on lifestyles and personal characteristics, as well as the collection of anthropometric measures and biological samples. Participants are tracked directly through follow-up assessments involving subsets of subjects, and indirectly through routine linkage with various administrative health databases including mortality, hospital episode statistics, cancer screening, and primary care. The residential data are available in the UKB database, including the exact dates participants moved to each residence and the related geocoded locations with 100m rounding. The data were validated internally, and the mobility history was continuously updated through general practitioner registration or direct reporting by the participants. The cohort profile was described in previous publications.[24,25] Details regarding the UKB database can be found on the showcase website (https://biobank.ndph.ox.ac.uk/showcase/).

Mortality outcomes

At the time of enrolment, the subjects consented access to a variety of personal information, including linked electronic health records and residential address location. The health data included, among other outcomes, mortality records routinely extracted and updated from the UK national registers. We defined mortality outcomes based on the date and primary cause of death included in the main dataset following the International Classification of Disease 10th (ICD-10) revision: specifically, we defined all-cause (ICD-10 codes: A00-U99), non-accidental (A00-R99), cardiovascular (I00-I99), respiratory (J00-J99), and lung cancer (C34) mortality outcomes.

Exposure assessment and linkage

Existing $PM_{2.5}$ exposure data available in the UKB database consist of annual average levels for the year 2010 assigned at the residential address established at recruitment using a Europe-wide land-use regression (LUR) model with a resolution of 100x100m.[26] This exposure linkage covered the subset of participants living within 400km of London (see the appendix). To enable assignment of time-varying exposures across the whole cohort, we used an improved exposure data source and the full residential history of each participant. The new exposure data source consisted of daily $PM_{2.5}$ levels predicted on a 1-km grid across the UK in the period 2003-2021 using a hybrid spatio-temporal machine learning (ML) model. The model used a random forest algorithm trained using ground monitor series and a series of spatial and spatio-temporal predictors, including outputs from emission-dispersion models, remote sensing satellite data, as well as land use and traffic variables, among others.[27] The model performance, assessed using monitor-based cross-validation, provided an overall coefficient of determination (R^2) of 0.77 at daily scale.

The exposure assignment process is described in previous work,[28] and briefly summarised here. The process was performed in two steps. First, we constructed daily exposure series for each home location by combining gridded exposure values using bilinear interpolation, a process that performs a weighted average of the four nearest 1x1km cells using weights computed as the inverse distance of the location from the corresponding centroids. This approach allowed maximising the exposure information while masking the original data, thus preventing back-tracing of individual locations. Second, we composed the subject-specific exposure profiles by linking the daily series for corresponding periods determined by the residential history. The daily subject-specific exposure data were aggregated in annual series corresponding to calendar years to facilitate the model fitting, as described below.

Study design and statistical analysis

We performed a time-to-event analysis based on an extended Cox proportional hazard model with time-varying predictors. Subjects were censored at the time of death, lost to follow-up, or the administrative end of follow-up (set here to 31/12/2021), whichever came first. Events were defined as death for any cause or specific causes only. The Cox model was specified using calendar time as the temporal axis and stratified by assessment centre, sex, and year of birth, thus ensuring a strong control for temporal trends and differential risks by age. The time-to-event data was split by calendar year (defining yearly subject/periods starting on the 1st of January) to assign exposure levels over a window of eight years, consistently with previous studies.[14] Specifically, for each sub-period, we assigned the yearly averaged $PM_{2.5}$ from the previous calendar year (defined here as lag 0) to eight years earlier (lag 7), so that exposure always preceded a mortality event. Given the need to define complete 8-year exposure histories and that the exposure data covered the period from 2003 on, the follow-up period effectively started on 01/01/2011 for all the subjects. Events that occurred before this date and subjects with incomplete exposure history were dropped from the analysis. We assumed a linear relationship and, in the main analysis, we defined the exposure index as the average of $PM_{2.5}$ within four windows: 1 year (lag 0), 2 years (lag 0-1), 5 years (lag 0-4) and 8 years (lag 0-7). Each window was investigated in separate models.

We specified six confounder models with increasing control for individual and area-level covariates measured at baseline, all entered as linear or categorical terms. The selection of confounders was based on a directed acyclic graph that defined the assumed causal pathways (Figure S1 in the appendix). Model 1 only included the matching variables used for stratification (assessment centre, sex, and year of birth). Models 2 and 3 added control for socio-economic status, using contextual and individual-level socio-economic variables, respectively. Specifically, Model 2 used the area-level Townsend deprivation index measured in 2010, while Model 3 included ethnic background, education level, household income, and employment status determined at recruitment. Model 4 featured all the variables in the previous models, while Model 5 also featured urban-rural classification and greenspace, two additional area-level variables rarely controlled for in published studies. Finally, Model 6 added a list of lifestyle-related factors rarely available in previous studies, including smoking status and intensity, alcohol intake, waist-to-hip ratio, physical activity (measured using the International Physical Activity Questionnaire (IPAQ) scale), and living alone. Details on the variable definitions are provided in the supplementary appendix (first section and Table S1).

In secondary analyses, we inspected the lag structure using distributed lag models (DLMs),[29] using the extended version for time-to-event data.[30] Specifically, we applied two different functions to specify the lag-response relationship in the fully-adjusted model (Model 6): natural cubic splines with 4 degrees of freedom, and strata defining steps for lags 0, 1-2, 3-4, and 5-7.

We performed several sensitivity analyses. First, we replicated the main model using alternative exposure sources and definitions, specifically using time-constant indices using either the value previously assigned from the LUR model or the average for 2010 from our ML model.[26,27] In addition, we repeated the analysis truncating the follow-up to the pre-2020 or the post-2013 periods. The former was used to remove the years affected by the COVID-19 pandemic, and the latter to include a washout period in order to account for potential healthy-volunteer and other selection biases, as recently recommended.[31]

Estimates of the associations were reported as hazard ratios (HRs) for all-cause and specific mortality outcomes per 10 $\mu g/m^3$ increments in $PM_{2.5}$, with 95% confidence intervals (CI). Missing values in the baseline covariates were imputed using multiple imputation by chained equation (MICE), producing five imputed datasets, with estimates combined using Rubin's rule.[32]

Results

Population and exposure characteristics

The original dataset included 502,381 individuals, from which 4,291 (0.85%) participants were excluded due to (partially) missing exposure histories, providing a final cohort of 498,090 individuals. The participants were followed up for an average of 10.3 years, with a total of 5,117,660 person-years. During the follow-up, there were 33,817 deaths among the selected participants, with 31,791 for non-accidental causes, 6,904 and 2,461 for cardiovascular and respiratory outcomes, respectively, and 2,820 for lung cancer. Tables 1-2 report descriptive statistics for the selected cohort. The average age at recruitment was 56.5 years. The participants included 54.5% of women, and it was overwhelmingly represented by people of white ethnicity (94.1%) and with high education (69.5% with a high school diploma or higher). Only 10.4% of the cohort were smokers at enrolment, most of them (65.0%) performed moderate/high physical activity, and only 8.1% consumed no alcohol. The majority (84.3%) lived in urban areas.

Figure 1 illustrates the box-and-whiskers plot with the distribution of the annual average exposure to $PM_{2.5}$ across the years. The plot shows that, across the study period, most of the participants were exposed to levels of air pollution that can be considered low (<15 μg/m³), and all lived in areas below the current UK air quality guideline level (25 μg/m³). Exposure levels show a noticeable temporal variation and a reduction from 2015 onwards, with most individuals living in areas with $PM_{2.5}$ levels below 10 μg/m³ in 2021. However, the majority of them were still exposed to concentrations above the current recommended limit by the World Health Organization (5 μg/m³).

Mortality risks associated with long-term PM2.5 exposure

The estimates of the association between long-term exposure to $PM_{2.5}$ and mortality are reported in Table 3, with combinations of causes of death, exposure windows, and various levels of covariate adjustments. In the fully-adjusted model (Model 6) and considering a cumulative exposure in the last eight years (lag 0-7), an increase of 10 μg/m³ in $PM_{2.5}$ was associated with an HR of 1.27 (95%CI: 1.06-1.53) and 1.24 (95%CI: 1.03-1.50) for all-cause and non-accidental mortality, respectively. There was evidence of positive associations also with respiratory (HR=2.07, 95%CI: 1.04-4.10) and lung cancer (HR=1.66, 95%CI: 0.86-3.19) mortality, although with wider confidence intervals, due to the smaller number of events. Surprisingly, cardiovascular mortality was not associated with long-term exposure to $PM_{2.5}$ (HR=0.88 95%CI: 0.59-1.31). In general, the associations were positive across lags and their magnitude increased with longer lag windows, with the exception of respiratory mortality.

The comparison of the estimates across models with different covariate adjustments in Table 3 indicates strong confounding effects, more pronounced for area-level factors compared with individual-level characteristics. The HR for non-accidental mortality and lag 0-7 dropped from 4.11 (95%CI: 3.50-4.82) in the basic model with only stratifying variables (Model 1) to 1.43 (95%CI: 1.21-1.69) when area-level deprivation was accounted for (Model 2). Adding individual-level socio-economic variables (Model 4) had little effect, while the inclusion of urban-rural residential settings and area-level greenspace (Model 5) further reduced the risk to an HR of 1.31 (95%CI: 1.09-1.59). Interestingly, the results suggested noticeable confounding effects from other individual-level covariates represented by lifestyle variables (Model 6) for most of the mortality causes, with the estimated risk decreasing further for all-cause, non-accidental, and lung cancer mortality, while remaining stable for respiratory causes.

Analysis of the lag structure

In secondary analyses, we investigated potential differences within the relevant exposure window by extending the Cox model with the simpler cumulative exposure through DLMs, specifically using natural splines and step functions. The results are reported in Figure 2, with the estimated lag-response relationships reporting the HR along lags for an increase of 10 μg/m³ in $PM_{2.5}$. It should be acknowledged immediately that the relatively low number of events and associated statistical power resulted in low precision and wide confidence intervals, preventing firm conclusions. However, the graphs provide interesting suggestions. For instance, the analysis of the lag-response curves for all-cause and non-accidental mortality indicates positive associations with exposures experienced in the last year (lag 0) and at longer timeframes (lag 5-7). In contrast, respiratory and lung cancer mortality showed associations for different exposure windows, specifically lag 0-2 for the former and lag 3-5 for the latter, suggesting differential temporal relationships related to different patho-physiological processes. The estimates of the overall cumulative associations are close to the main model with a simple exposure definition assuming a constant effect across lags (Table S2 in the appendix).

Sensitivity analyses using different exposure indices and follow-up periods

In the first sensitivity analysis, we replicated the main model using alternative exposure summaries and sources. The assessment included a subset of 457,925 UKB participants (91.9% of the cohort used in the main analysis) with linked $PM_{2.5}$ levels assigned from the original LUR model and the new ML model.[26,27] The correlation between the two sources for the year 2010 is low, with a Pearson *r* coefficient of 0.25. The analysis of the $PM_{2.5}$ distributions, shown in Figure S2 (appendix), reveals a peculiar pattern in the original data from the LUR model, with the $PM_{2.5}$ left-truncated at about 8 μg/m³, while the new ML model provides a wider exposure range and bimodal shape probably due to urban/rural residential locations.

Figure 3 illustrates the comparison of the HRs estimated for each mortality cause using the original time-varying annual $PM_{2.5}$ from the ML model, the fixed annual average in 2010 from the same model, and the corresponding value for the same year from the existing UKB data derived from the LUR model (see Table S3 in the appendix for the actual HR figures). The graph shows that the use of time-varying exposure summaries results in generally higher HRs, although with wider confidence intervals.

Further sensitivity analyses assessed the sensitivity to the right and left truncation of the follow-up period, with results reported in Table S4 (appendix). The exclusion of the period affected by the COVID-19 pandemic (pre-2020) generated very similar results. In contrast, the inclusion of a washout period with the restriction of the follow-up to the post-2013 years resulted in a general increase of the HR for all the mortality causes.

Discussion

This epidemiological study assessed long-term associations between time-varying $PM_{2.5}$ exposure and mortality in the UK Biobank cohort. Results show increased risks of death for all causes as well as non-accidental, respiratory, and lung cancer, but not for cardiovascular outcomes. The comparison of models with different levels of adjustment for confounders indicates the important role of contextual variables such as area-level deprivation and greenspace, but also the need to control for individual-level factors related to lifestyle characteristics. The analysis features the innovative use of statistical techniques to analyse relevant lag structures through the implementation of distributed lag models within Cox proportional hazard regression, with suggestions of possible differences in the relevant exposure windows across mortality outcomes.

Our risk associations are generally higher than those reported in meta-analyses as well as large-scale or pooled studies, both for non-accidental mortality and other causes, especially respiratory and lung cancer.[10,11,13-17] Prior results for the UK are sparser and less consistent: an analysis of a national English cohort of patients registered to clinical practices reported an HR of 1.13 (95% CI: 1.00-1.29) for all-cause mortality,[21] while a recent analysis of the UKB cohort found a higher risk of 1.27 (95% CI: 1.05-1.55),[23] similar to that found by our study for all-cause mortality, although using time-constant exposure summaries and a less strict confounding control compared to our analysis. There can be multiple explanations for such differences. First, the use of more informative exposure data can have led to a reduction in exposure misclassification and a better ability to identify risk associations, as discussed below. Second, the risk could be steeper at the relatively low exposure ranges of the UKB cohort, as previously reported.[16,17,33] Interestingly, the higher size of the risk estimates using time-varying exposures and the findings about the confounding effects of lifestyle and other individual-level factors are consistent with other published analyses.[20,34]

In contrast, we found no evidence of an association with cardiovascular mortality. Such a link is in fact well established in the epidemiological literature summarised above, although conflicting results were reported from the UK.[21-23] A potential explanation is the relatively young age of the UKB participants, given the specific recruitment constraints (40-69 years of age) and the still relatively short follow-up, although this issue should be studied further in future analyses. In general, for all outcomes, differences can be due to variations in population susceptibility and specific PM composition. It must also be noted that, in any case, the relatively wide confidence intervals include ranges more consistent with the literature.

Important strengths of this study are the fine control for confounding and the use of time-varying exposure information. Most of the studies in the literature are based on large administrative cohorts with limited individual-level information,[13,14,17] or the pooling of small and heterogeneous cohorts,[16,22] although with some exceptions.[34] The UKB database offers a unique combination where information on both contextual and individual-level characteristics was collected for a large sample of half a million participants. The results reported in Table 3 demonstrate the potential biases arising from incomplete confounding control, with area-level deprivation, rural-urban settings, and greenspace, as well as individual lifestyle factors, all featuring as important predictors and potential confounders. Similarly, the analysis takes advantage of time-varying exposure information, provided by the detailed reconstruction of individual-level daily exposure histories to $PM_{2.5}$ for the full cohort using state-of-the-art spatio-temporal machine learning models.[27] With some exceptions,[14,35] most of the published studies used simpler effect summaries, based for instance on time-invariant exposure indices (for instance the average $PM_{2.5}$ in a given year),[16,17] or alternatively using time-varying measures represented only by the current year's level.[13] Previous methodological works highlighted the benefit of using time-varying exposure summaries based on longer exposure windows and reconstructed using exposure models with high spatial and temporal resolution.[20,36] This evidence is consistent with the results of our sensitivity analysis in Figure 3.

An original aspect of this study is the application of flexible methods to assess the temporal structure of the association and identify relevant exposure windows across different outcomes. This issue has been addressed in the large epidemiological literature for other risk factors such as smoking,[37] radon,[38] or medications,[39] but apart from some exceptions,[40,41] it has been notably overlooked in the research on the long-term risks of air pollution. In particular, the application of DLMs offers an easily adaptable framework where different functions can be applied to model the lag-response relationship, and it takes full advantage of finely temporally disaggregated exposure histories. The DLM methodology was originally implemented and is commonly used in time series studies to analyse short-term risks of environmental factors.[29,42,43] More recently, it has been generalised beyond time series data and it is

now applicable for the analysis of long-term associations,[30] although with limited applications so far.[41] Our study demonstrates its potential, and although the large uncertainty in the results prevents drawing firm conclusions in this specific case, it can be replicated in future follow-up updates of the UKB cohort and other studies.

An important limitation of the UKB cohort is its voluntary participation with a low response rate of 5.5%,[44] resulting in a selection of participants non-representative of the UK population.[45] While it has been established that representativeness is not a requirement for valid epidemiological inference,[46,47] a strong selection can result in biased estimates in the presence of unmeasured risk factors that act as colliders when associated with both the exposure and the probability of selection.[31,47,48] While we cannot verify this hypothesis, we included a washout period in a sensitivity analysis, recommended in a recent study as a measure to attenuate the bias.[31] This resulted in higher effect estimates, indicating that in case this bias occurred, our estimates would be conservative.

Other limitations must be acknowledged. First, despite the size of the UKB cohort, its follow-up is still relatively short and results in a small number of health events compared to other published studies. This affected the precision of the estimates, in particular for cause-specific mortality outcomes, and prevented supplementary analysis such as on the shape of the exposure-response curve and potential non-linearities. In addition, most of the potential confounders included in the analysis were only measured at baseline, and we could not account for potential changes during the follow-up, in particular related to lifestyle factors. Regarding the exposure measurement, while we relied on high-resolution $PM_{2.5}$ levels derived from the spatio-temporal ML model, these only represent residential concentrations, which can result in some degree of exposure misclassification. Furthermore, the exposure reconstruction could only cover the period from 2003 on, thus limiting the temporal windows we could assess in our analysis. Finally, we were not able to adjust for co-pollutants such as nitrogen dioxide (NO_2) or ozone (O_3). However, the linkage of these and other environmental stressors to the UKB database is ongoing and this limitation can be addressed in future analyses.

In conclusion, this epidemiological assessment examined the association between long-term exposure to $PM_{2.5}$ and mortality in the UKB cohort, a large prospective study of more than half a million British adults, using finely-reconstructed exposures, strict control for confounding, and advanced study designs and statistical methods. The results indicate significant risks associated with $PM_{2.5}$ for all-cause, non-accidental, and respiratory causes, as well as lung cancer, but not for cardiovascular outcomes.

Funding: This work was supported by the Nagasaki University "Doctoral Program for World-leading Innovative and Smart Education" for Global Health (WISE), KENKYU SHIDO KEIHI ("the Research Grant") and KYOIKU KENKYU SHIEN KEIHI ("the Stipend"), and by the Medical Research Council UK (Grant ID: MR/Y003330/1).

Competing financial interests: The author declares he has no actual or potential conflict of interest.

Data and code: The main database of the UK Biobank can be accessed via an application through the UK Biobank online Access Management System (AMS) (https://www.ukbiobank.ac.uk/). The PM2.5 exposure was expressly linked in this analysis and will be fully linked to the main database and released six months after publication. The full R code to prepare the data and perform the analysis is available upon request to the last author.

Acknowledgements:) This research has been conducted using the UK Biobank Resource under application number 56431. This work was part of a PhD project supported by the Nagasaki University WISE Programme.

Tables

Table 1. Descriptive statistics (mean with 5th-95th percentile range and missing) for continuous baseline characteristics[§] in the selected UKB cohort (498,090 participants recruited in the period 2006-2010).

Age at baseline	Years	**56.50 (42.00 to 68.00)**
	Missing (%)	0 (0.0%)
Waist-to-hip ratio	-	0.87 (0.73 to 1.02)
	Missing (%)	2,165 (0.4%)
Smoking intensity	Pack-years	8.22 (0.00 to 40.62)
	Missing (%)	77,321 (15.5%)
Townsend deprivation index	-	-1.30 (-5.05 to 4.92)
	Missing (%)	616 (0.1%)
Greenspace	%	45.01 (15.58 to 86.94)
	Missing (%)	60,318 (12.1%)

[§]See the text and Table S1 in the appendix for definitions and additional information.

Table 2. Distributions of categorical baseline characteristics[§] in the selected UKB cohort (498,090 participants recruited in the period 2006-2010). See the text and Table S1 in the appendix for definitions and additional information.

Sex	Female	271,660 (54.5%)	Rural/urban classification	Urban	420,043 (84.3%)
	Male	226,430 (45.5%)		Town/fringe	37,256 (7.5%)
	Missing (%)	0 (0.0%)		Village/Rural	35,893 (7.2%)
Ethnic background	White	468,508 (94.1%)		Missing (%)	4,898 (1.0%)
	Other	26,847 (5.4%)	Physical activity	Low	75,255 (15.1%)
	Missing (%)	2,735 (0.5%)		Moderate	162,774 (32.7%)
Employment	Employed	285,717 (57.4%)		High	161,041 (32.3%)
	Retired	164,864 (33.1%)		Missing (%)	99,020 (19.9%)
	Other	41,926 (8.4%)	Alcohol consumption	Never	40,046 (8.0%)
	Missing (%)	5,583 (1.1%)		Special occasions only	57,388 (11.5%)
Education	Low	83,964 (16.9%)		One to three times a month	55,424 (11.1%)
	Professional Qualification	57,948 (11.6%)		Once or twice a week	128,297 (25.8%)
	Highschool diploma	186,136 (37.4%)		Three or four times a week	114,678 (23.0%)
	College/University degree	160,098 (32.1%)		Daily or almost daily	100,795 (20.2%)
	Missing (%)	9,944 (2.0%)		Missing (%)	1,462 (0.3%)
Income (£)	Less than 18,000	95,735 (19.2%)	Smoking status	Never	271,865 (54.6%)
	18,000 to 30,999	107,266 (21.5%)		Previous	171,301 (34.4%)
	31,000 to 51,999	110,145 (22.1%)		Current	52,037 (10.4%)
	52,000 to 100,000	85,892 (17.2%)		Missing (%)	2,887 (0.6%)
	Greater than 100,000	22,831 (4.6%)	Living alone	No	402,043 (80.7%)
	Missing (%)	76,221 (15.3%)		Yes	92,937 (18.7%)
				Missing (%)	3,110 (0.6%)

[§]See the text and Table S1 in the appendix for definitions and additional information.

Table 3. Hazard ratios (HRs, with 95% confidence intervals) of different mortality causes associated with an increase of 10 μg/m³ in $PM_{2.5}$ in the UKB cohort, for combinations of length of exposure windows and confounding control.

	Lag	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
All causes (33,817 events)	0	3.12 (2.73-3.56)	1.37 (1.19-1.56)	2.20 (1.92-2.51)	1.33 (1.16-1.53)	1.25 (1.07-1.45)	1.20 (1.03-1.39)
	0-1	3.49 (3.04-4.01)	1.40 (1.22-1.62)	2.38 (2.07-2.73)	1.37 (1.19-1.58)	1.27 (1.09-1.50)	1.21 (1.03-1.43)
	0-4	3.96 (3.42-4.58)	1.44 (1.24-1.67)	2.60 (2.24-3.02)	1.40 (1.21-1.63)	1.30 (1.10-1.55)	1.24 (1.04-1.47)
	0-7	4.44 (3.80-5.18)	1.47 (1.26-1.73)	2.84 (2.43-3.33)	1.45 (1.23-1.70)	1.34 (1.12-1.61)	1.27 (1.06-1.53)
Non-accidental (31,791 events)	0	2.95 (2.58-3.39)	1.34 (1.17-1.54)	2.11 (1.84-2.43)	1.32 (1.14-1.51)	1.24 (1.06-1.44)	1.18 (1.01-1.38)
	0-1	3.26 (2.82-3.75)	1.37 (1.18-1.58)	2.26 (1.95-2.61)	1.34 (1.16-1.55)	1.25 (1.06-1.48)	1.19 (1.01-1.40)
	0-4	3.66 (3.15-4.26)	1.39 (1.19-1.62)	2.45 (2.10-2.85)	1.36 (1.17-1.59)	1.27 (1.06-1.51)	1.20 (1.01-1.43)
	0-7	4.11 (3.50-4.82)	1.43 (1.21-1.69)	2.68 (2.28-3.15)	1.41 (1.20-1.67)	1.31 (1.09-1.59)	1.24 (1.03-1.50)
Cardiovascular (6,904 events)	0	3.59 (2.67-4.83)	1.15 (0.86-1.55)	2.08 (1.54-2.80)	1.07 (0.79-1.44)	1.05 (0.76-1.47)	1.02 (0.73-1.41)
	0-1	3.99 (2.93-5.43)	1.13 (0.83-1.54)	2.19 (1.60-2.99)	1.04 (0.76-1.42)	1.02 (0.72-1.45)	0.97 (0.69-1.38)
	0-4	4.31 (3.12-5.96)	1.06 (0.77-1.47)	2.25 (1.62-3.12)	0.98 (0.70-1.36)	0.94 (0.64-1.36)	0.90 (0.62-1.30)
	0-7	4.80 (3.40-6.77)	1.04 (0.73-1.48)	2.41 (1.70-3.42)	0.97 (0.68-1.38)	0.92 (0.62-1.38)	0.88 (0.59-1.31)
Respiratory (2,461 events)	0	12.23 (7.33-20.41)	2.01 (1.21-3.37)	5.67 (3.36-9.57)	1.93 (1.14-3.26)	2.25 (1.24-4.09)	2.18 (1.21-3.94)
	0-1	14.34 (8.48-24.23)	2.04 (1.20-3.46)	6.28 (3.67-10.76)	1.95 (1.14-3.36)	2.34 (1.25-4.37)	2.23 (1.20-4.16)
	0-4	17.09 (9.87-29.58)	2.05 (1.17-3.57)	7.00 (3.99-12.28)	1.96 (1.11-3.46)	2.36 (1.22-4.56)	2.23 (1.16-4.28)
	0-7	18.63 (10.45-33.21)	1.93 (1.08-3.47)	7.28 (4.04-13.11)	1.88 (1.03-3.41)	2.21 (1.11-4.40)	2.07 (1.04-4.10)
Lung cancer (2,820 events)	0	8.27 (5.20-13.15)	1.72 (1.08-2.74)	5.01 (3.11-8.05)	1.80 (1.12-2.89)	1.77 (1.05-3.01)	1.52 (0.90-2.58)
	0-1	9.39 (5.80-15.20)	1.67 (1.03-2.72)	5.40 (3.29-8.84)	1.74 (1.07-2.86)	1.71 (0.98-2.99)	1.42 (0.81-2.49)
	0-4	13.13 (7.86-21.93)	1.90 (1.13-3.20)	7.13 (4.21-12.07)	2.01 (1.18-3.41)	2.02 (1.10-3.70)	1.62 (0.89-2.98)
	0-7	15.96 (9.20-27.69)	1.91 (1.09-3.35)	8.32 (4.73-14.63)	2.07 (1.17-3.66)	2.06 (1.08-3.95)	1.66 (0.86-3.19)

Model 1: stratifying variables only (assessment centre, sex, and month of age)

Model2: Model 1 + area-level deprivation (Townsend deprivation index, TDI)

Model 3: Model 1 + individual-level socio-economic variables (ethnic background, education level, household income, and employment status)

Model 4: Model 2 + Model 3

Model 5: Model 4 + urban/rural classification and greenspace

Model 6: Model 5 + individual-level lifestyle factors (smoking status and pack/years, alcohol intake, physical activity, waist-to-hip ratio, living alone)

Figures

Figure 1. Distribution of annual average exposure to $PM_{2.5}$ across participants of the UKB cohort in the period 2003-2021, with limits corresponding to air quality guidelines/directives (AQG and AQD) from the World Health Organization (WHO), European Union (EU), and the United Kingdom (UK).

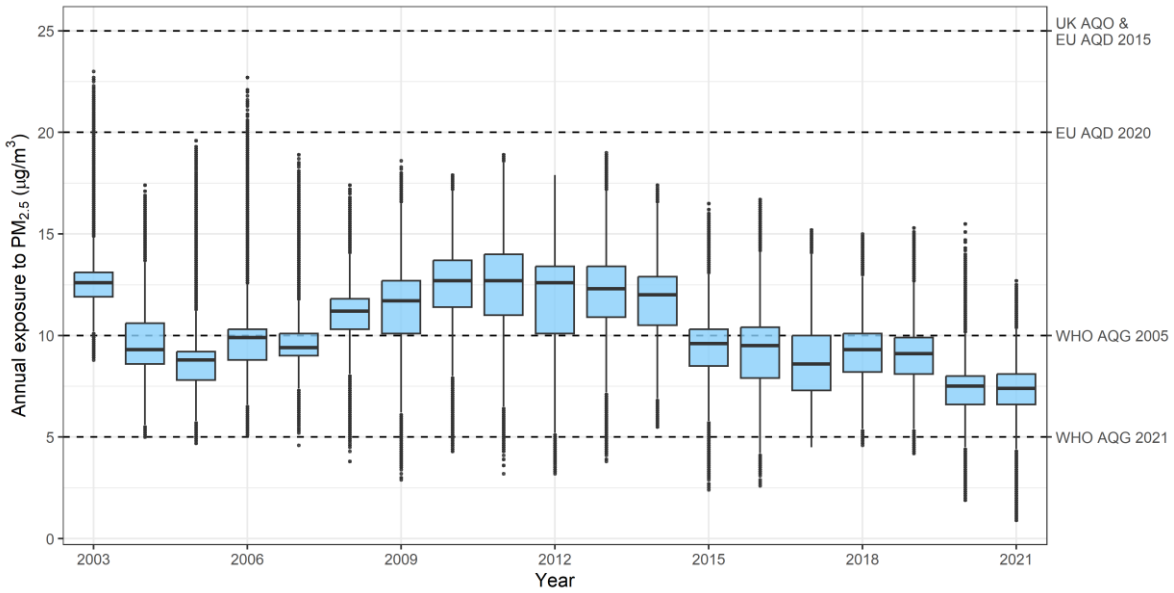


Figure 2. Lag-response relationships representing the hazard ratios (HRs, with 95% confidence intervals) of different mortality causes associated with an increase of 10 µg/m³ in $PM_{2.5}$ along an 8-year exposure window (lag 0-7), estimated by distributed lag models (DLM) with splines (line with 95% confidence intervals area) and step functions (points with 95% confidence intervals bars), respectively, using the fully-adjusted confounder model (Model 6 in Table 3).

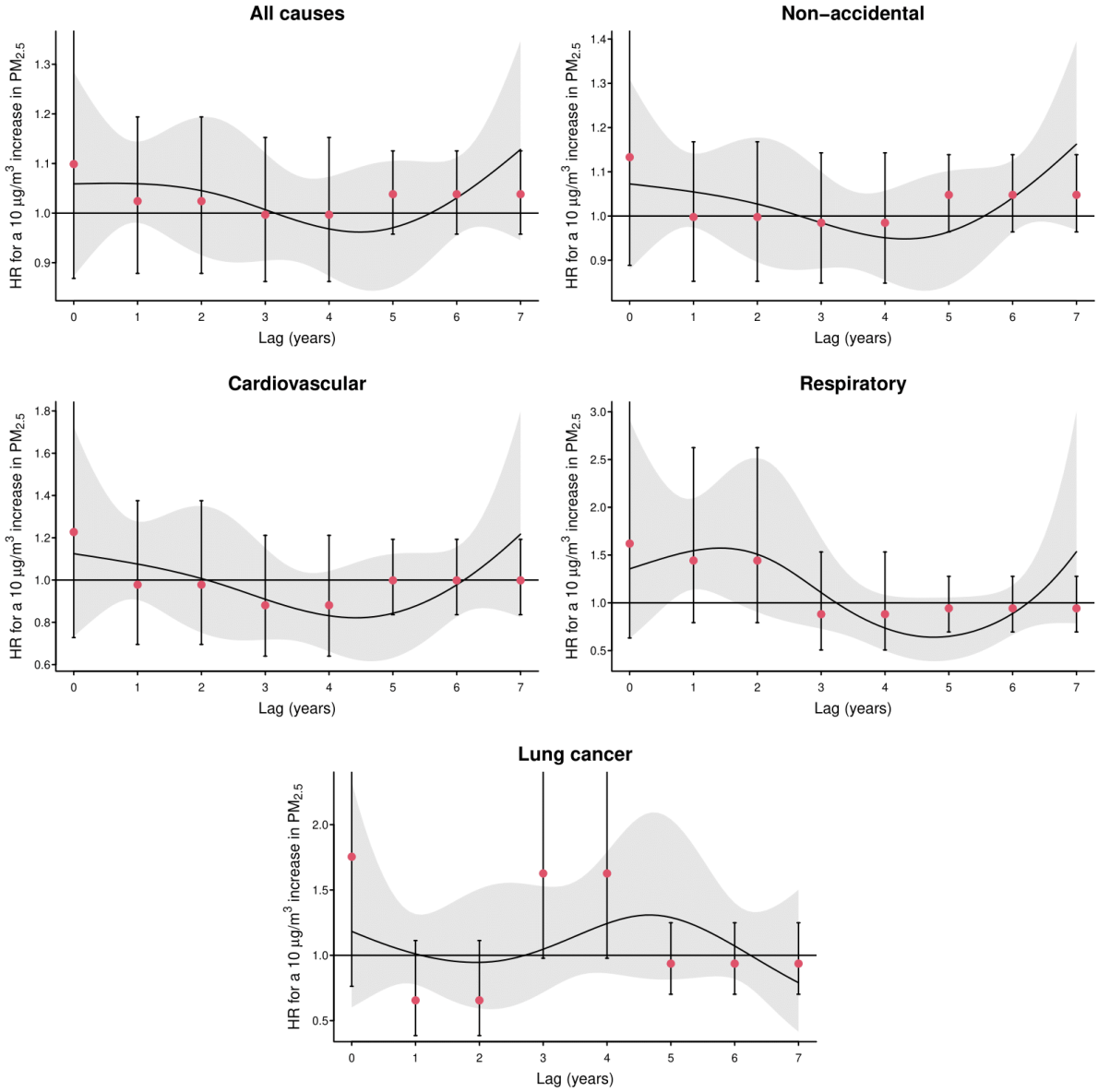
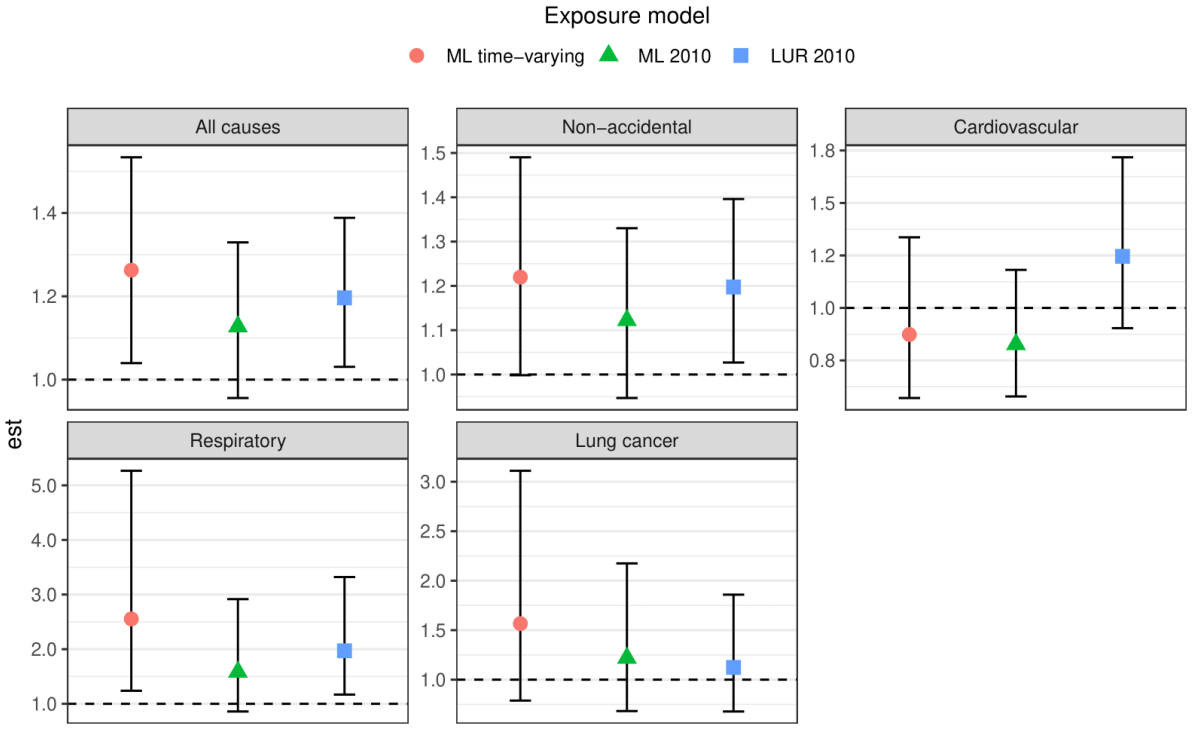


Figure 3. Comparison of hazard ratios (HRs, with 95% confidence intervals) of different mortality causes associated with an increase of 10 μg/m³ in $PM_{2.5}$, estimated using alternative exposure sources and indices, using the fully-adjusted confounder model (Model 6 in Table 3).



References

1. Burnett R, Chen H, Szyszkowicz M, et al. Global estimates of mortality associated with long-term exposure to outdoor fine particulate matter. *Proceedings of the National Academy of Sciences*. 2018;115(38):9592-9597.
2. Wu X, Braun D, Schwartz J, Kioumourtzoglou MA, Dominici F. Evaluating the impact of long-term exposure to fine particulate matter on mortality among the elderly. *Science Advances*. 2020;6(29).
3. Hoffmann B, Brunekreef B, Andersen ZJ, Forastiere F, Boogaard H. Benefits of future clean air policies in Europe. *Environmental Epidemiology*. 2022;6(5):e221.
4. Rajagopalan S, Al-Kindi SG, Brook RD. Air pollution and cardiovascular disease: JACC state-of-the-art review. *Journal of the American College of Cardiology*. 2018;72(17):2054-2070.
5. Doiron D, de Hoogh K, Probst-Hensch N, et al. Air pollution, lung function and COPD: results from the population-based UK Biobank study. *European Respiratory Journal*. 2019;54(1).
6. Hamra GB, Guha N, Cohen A, et al. Outdoor particulate matter exposure and lung cancer: a systematic review and meta-analysis. *Environmental Health Perspectives*. 2014;122(9):906-911.
7. Afsar B, Afsar RE, Kanbay A, Covic A, Ortiz A, Kanbay M. Air pollution and kidney disease: review of current evidence. *Clinical Kidney Journal*. 2018;12(1):19-32.
8. Fu P, Guo X, Cheung FMH, Yung KKL. The association between PM2.5 exposure and neurological disorders: a systematic review and meta-analysis. *Science of The Total Environment*. 2019;655:1240-1248.
9. Yang B-Y, Fan S, Thiering E, et al. Ambient air pollution and diabetes: a systematic review and meta-analysis. *Environmental Research*. 2020;180:108817.
10. Chen J, Hoek G. Long-term exposure to PM and all-cause and cause-specific mortality: A systematic review and meta-analysis. *Environment International*. 2020;143:105974.
11. Pope CA, Coleman N, Pond ZA, Burnett RT. Fine particulate air pollution and human mortality: 25+ years of cohort studies. *Environmental Research*. 2020;183:108924.
12. Krewski D, Burnett R, Goldberg M, et al. Overview of the reanalysis of the Harvard Six Cities Study and American Cancer Society Study of Particulate Air Pollution and Mortality. *Journal of Toxicology and Environmental Health, Part A*. 2003;66(16-19):1507-1552.
13. Di Q, Wang Y, Zanobetti A, et al. Air pollution and mortality in the Medicare population. *New England Journal of Medicine*. 2017;376(26):2513-2522.
14. Crouse DL, Peters PA, Hystad P, et al. Ambient PM2.5, O3, and NO2 exposures and associations with mortality over 16 years of follow-up in the Canadian Census Health and Environment Cohort (CanCHEC). *Environmental Health Perspectives*. 2015;123(11):1180-1186.
15. Beelen R, Raaschou-Nielsen O, Stafoggia M, et al. Effects of long-term exposure to air pollution on natural-cause mortality: an analysis of 22 European cohorts within the multicentre ESCAPE project. *The Lancet*. 2014;383(9919):785-795.
16. Strak M, Weinmayr G, Rodopoulou S, et al. Long term exposure to low level air pollution and mortality in eight European cohorts within the ELAPSE project: pooled analysis. *British Medical Journa*. 2021;374:n1904.
17. Stafoggia M, Oftedal B, Chen J, et al. Long-term exposure to low ambient air pollution concentrations and mortality among 28 million people: results from seven large European cohorts within the ELAPSE project. *The Lancet Planetary Health*. 2022;6(1):e9-e18.
18. de Hoogh K, Chen J, Gulliver J, et al. Spatial PM2.5, NO2, O3 and BC models for Western Europe - Evaluation of spatiotemporal stability. *Environment International*. 2018;120:81-92.
19. van Donkelaar A, Martin RV, Li C, Burnett RT. Regional estimates of chemical composition of fine particulate matter using a combined geoscience-statistical method with information from satellites, models, and monitors. *Environmental Science \& Technology*. 2019;53(5):2595–2611.

20. Crouse DL, Erickson AC, Christidis T, et al. Evaluating the sensitivity of PM2.5-mortality associations to the spatial and temporal scale of exposure assessment. *Epidemiology*. 2020;31(2):168-176.
21. Carey IM, Atkinson RW, Kent AJ, van Staa T, Cook DG, Anderson HR. Mortality associations with long-term exposure to outdoor air pollution in a national English cohort. *American Journal of Respiratory and Critical Care Medicine*. 2013;187(11):1226-1233.
22. Dehbi H-M, Blangiardo M, Gulliver J, et al. Air pollution and cardiovascular mortality with over 25years follow-up: a combined analysis of two British cohorts. *Environment International*. 2017;99:275-281.
23. Wang M, Zhou T, Song Q, et al. Ambient air pollution, healthy diet and vegetable intakes, and mortality: a prospective UK Biobank study. *International Journal of Epidemiology*. 2022;51(4):1243-1253.
24. Sudlow C, Gallacher J, Allen N, et al. UK Biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Medicine*. 2015;12(3):e1001779.
25. Conroy M, Sellors J, Effingham M, et al. The advantages of UK Biobank's open-access strategy for health research. *Journal of Internal Medicine*. 2019;286(4):389-397.
26. Eeftens M, Beelen R, de Hoogh K, et al. 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. *Environmental Science \& Technology*. 2012;46(20):11195-11205.
27. Schneider R, Vicedo-Cabrera AM, Sera F, et al. A satellite-based spatio-temporal machine learning model to reconstruct daily PM2.5 concentrations across Great Britain. *Remote Sensing*. 2020;12(22):3803.
28. Vanoli J, Mistry MN, De La Cruz Libardi A, et al. Reconstructing individual-level exposures in cohort analyses of environmental risks: an example with the UK Biobank. *Journal of Exposure Science \& Environmental Epidemiology*. 2024:Online ahead of print. DOI: 10.1038/s41370-023-00635-w.
29. Gasparrini A, Armstrong B, Kenward MG. Distributed lag non-linear models. *Statistics in Medicine*. 2010;29(21):2224-2234.
30. Gasparrini A. Modeling exposure-lag-response associations with distributed lag non-linear models. *Statistics in Medicine*. 2014;33(5):881-899.
31. Chen C, Chen H, Kaufman JS, Benmarhnia T. Differential participation, a potential cause of spurious associations in observational cohorts in environmental epidemiology. *Epidemiology*. 2024;35(2):174–-184.
32. Marshall A, Altman DG, Holder RL, Royston P. Combining estimates of interest in prognostic modelling studies after multiple imputation: current practice and guidelines. *BMC Medical Research Methodology*. 2009;9(1).
33. Pappin AJ, Christidis T, Pinault LL, et al. Examining the shape of the association between low levels of fine particulate matter and mortality across three cycles of the Canadian Census Health and Environment Cohort. *Environmental Health Perspectives*. 2019;127(10).
34. Pinault L, Tjepkema M, Crouse DL, et al. Risk estimates of mortality attributed to low concentrations of ambient fine particulate matter in the Canadian community health survey cohort. *Environmental Health*. 2016;15(1):18.
35. Chen J, Braun D, Christidis T, et al. Long-Term exposure to low-level PM2.5 and mortality: investigation of heterogeneity by harmonizing analyses in large cohort studies in Canada, United States, and Europe. *Environmental Health Perspectives*. 2023;131(12).
36. Lefler JS, Higbee JD, Burnett RT, et al. Air pollution and mortality in a large, representative U.S. cohort: multiple-pollutant analyses, and spatial and temporal decompositions. *Environmental Health*. 2019;18(1):101.

37. Richardson DB, Terschuren C, Pohlabeln H, Jockel KH, Hoffmann W. Temporal patterns of association between cigarette smoking and leukemia risk. *Cancer Causes Control*. 2008;19(1):43-50.
38. Langholz B, Thomas D, Xiang A, Stram D. Latency analysis in epidemiologic studies of occupational exposures: application to the Colorado Plateau uranium miners cohort. *American Journal of Industrial Medicine*. 1999;35(3):246-256.
39. Abrahamowicz M, Bartlett G, Tamblyn R, du Berger R. Modeling cumulative dose and exposure duration provided insights regarding the associations between benzodiazepines and injuries. *Journal of Clinical Epidemiology*. 2006;59(4):393-403.
40. Schwartz J, Coull B, Laden F, Ryan L. The effect of dose and timing of dose on the association between airborne particles and survival. *Environmental Health Perspectives*. 2008;116(1):64-69.
41. Kriit HK, Andersson EM, Carlsen HK, et al. Using distributed lag non-linear models to estimate exposure lag-response associations between long-term air pollution exposure and incidence of cardiovascular disease. *International Journal of Environmental Research and Public Health*. 2022;19(5):2630.
42. Schwartz J. The distributed lag between air pollution and daily deaths. *Epidemiology*. 2000;11(3):320-326.
43. Armstrong B. Models for the relationship between ambient temperature and daily mortality. *Epidemiology*. 2006;17(6):624-631.
44. Swanson JM. The UK Biobank and selection bias. *The Lancet*. 2012;380(9837):110.
45. Fry A, Littlejohns TJ, Sudlow C, et al. Comparison of sociodemographic and health-related characteristics of UK Biobank participants with those of the general population. *American Journal of Epidemiology*. 2017;186(9):1026–1034.
46. Rothman KJ, Gallacher JE, Hatch EE. Why representativeness should be avoided. *International Journal of Epidemiology*. 2013;42(4):1012—-1014.
47. Richiardi L, Pizzi C, Pearce N. Commentary: Representativeness is usually not necessary and often should be avoided. *International Journal of Epidemiology*. 2013;42(4):1018—-1022.
48. Pizzi C, De Stavola B, Merletti F, et al. Sample selection and validity of exposure-disease association estimates in cohort studies. *Journal of Epidemiology \& Community Health*. 2010;65(5):407—-411.

Long-term associations between time-varying exposure to ambient $PM_{2.5}$ and mortality risks: an analysis of the UK Biobank

Appendix

Variable definitions and sources in the UKB database

The data used in the analysis were retrieved from the UKB database. The information was mostly collected at recruitment during the baseline assessment, with the exception of the follow-up and mortality data that are linked routinely and constantly updated. The other source of data is the time-varying $PM_{2.5}$ exposure that was generated by the spatio-temporal machine learning (ML) model (Ref 27 in the main text), which was linked separately and will be integrated into the UKB database soon. Table S1 below provides information on the selected variables. More details are available in the UKB Showcase (https://biobank.ndph.ox.ac.uk/showcase/), which includes a 'search' tab to gather information on each variable using the related Field IDs.

Table S1. Definition of variables in the UKB database used in the analysis, together with the corresponding Field IDs in the UKB database.

Variable	Field IDs	Notes
Date of birth	34, 52	Derived from year (Field 34) calendar month (Field 52) of birth, and assigned to the 15th day (the exact day is not provided)
Date of death	40000	Acquired from the central registry
Cause of death	40001	Underlying (primary) cause of death, coded using ICD-10
Date lost to follow-up		Date from which a participant is believed to be lost to follow-up (not updated from 2017)
Date of start of follow-up	53	Date when a participant attended a UKB assessment centre
Assessment centre	54	One of the 22 assessment centres at which participants were recruited
Sex	21022	Genetic sex
Age at baseline	31, 22001	Derived based on the date of birth and date of attending the initial assessment centre
Townsend deprivation index	22189	Assigned corresponding to the output area in which the participant's residence postcode is located
Ethnic background	21000	Amalgam of sequential branching questions
Employment	6142, 20119	Field 20119 was used to correct the original Field 6142
Education	6138	Education level based on qualifications
Income	738	Average total household income before tax
Waist-to-hip ratio	48, 49	Ratio of circumferences of waist (Field 48) and hip (Field 49)
Physical activity	22032	IPAQ score
Alcohol drinking status	1558	Based on the frequency of alcohol drinking
Smoking status	20116	Current/past smoking status of the participant
Smoking intensity	20161	Pack years of smoking (derived by combining various Fields)
Living alone	709, 670	Derived from information on the number of people in the household (Field 709) and type of accommodation (Field 670)
Urban-rural classification	20118	Derived by linking each participant's home postcode to the 2001 census from the Office of National Statistics
Greenspace	24500	Percentage of land-use type classified as greenspace in a buffer of 1000m around the participant's residence
$PM_{2.5}$ from LUR model	24006	$PM_{2.5}$ level ($\mu g/m^3$) at each participant's residence in 2010 derived from a Europe-wide land-use regression (LUR) model

Additional Supplementary Tables

Table S2. Hazard ratios (HRs, with 95% confidence intervals) of different mortality causes associated with an increase of 10 μg/m³ in PM2.5 in the UKB cohort, estimated using different models for modelling the lag-response relationships along an 8-year exposure window (lag 0-7) and the fully-adjusted confounder model (Model 6 in Table 3).

	Main model (lag 0-7 average)	Spline-DLM	Strata-DLM
All causes	1.27 (1.06-1.53)	1.29 (1.07-1.55)	1.28 (1.06-1.54)
Non-accidental	1.24 (1.03-1.50)	1.27 (1.05-1.54)	1.26 (1.04-1.52)
Cardiovascular	0.88 (0.59-1.31)	0.92 (0.62-1.38)	0.91 (0.60-1.36)
Respiratory	2.07 (1.04-4.10)	2.28 (1.14-4.57)	2.19 (1.09-4.38)
Lung cancer	1.66 (0.86-3.19)	1.61 (0.83-3.12)	1.64 (0.84-3.17)

Main model: corresponding to a simple DLM with a single stratum modelling a constant lag-response risk

Spline-DLM: using natural cubic splines with 4 degrees of freedom (two-equally-space internal knots and an intercept)

Strata-DLM: using strata defining steps for lags 0, 1-2, 3-4, and 5-7

Table S3. Comparison of the number of events for different mortality causes and corresponding hazard ratios (HRs, with 95% confidence intervals) associated with an increase of 10 μg/m³ in $PM_{2.5}$, estimated using alternative exposure sources and indices (see Figure 3 in the main text), using the fully-adjusted confounder model (Model 6 in Table 3).

	N	ML time-varying	ML 2010	LUR 2010
All causes	30,750	1.26 (1.04-1.53)	1.13 (0.96-1.33)	1.20 (1.03-1.39)
Non-accidental	28,934	1.22 (1.00-1.49)	1.12 (0.95-1.33)	1.20 (1.03-1.40)
Cardiovascular	6,287	0.87 (0.57-1.34)	0.83 (0.58-1.18)	1.25 (0.90-1.72)
Respiratory	2,262	2.55 (1.24-5.27)	1.58 (0.86-2.92)	1.97 (1.17-3.32)
Lung cancer	2,543	1.57 (0.79-3.11)	1.22 (0.68-2.17)	1.12 (0.68-1.86)

ML-time-varying: annual $PM_{2.5}$ varying during the follow-up period from a spatio-temporal machine learning model (Ref 23 in the main text)

ML-2010: annual $PM_{2.5}$ in 2010 from the model above.

LUR 2010: annual $PM_{2.5}$ in 2010 from a Europe-wide land-use regression (LUR) model (Ref 27 in the main text)

Table S4. Number of events for different mortality causes and corresponding hazard ratios (HRs, with 95% confidence intervals) associated with an increase of 10 μg/m³ in $PM_{2.5}$, estimated by truncating the follow-up to the pre-2020 (left) and post-2013 (right) periods, using the fully-adjusted confounder model (Model 6 in Table 3).

	N	Pre-2020 period	N	Post-2013 period
All causes	29,752	1.28 (1.05-1.56)	26,180	1.33 (1.09-1.62)
Non-accidental	28,238	1.24 (1.02-1.52)	24,423	1.30 (1.06-1.58)
Cardiovascular	6,029	0.85 (0.55-1.30)	5,376	0.99 (0.65-1.52)
Respiratory	2,203	2.22 (1.08-4.56)	2,046	2.38 (1.16-4.89)
Lung cancer	2,578	1.86 (0.94-3.70)	2,002	2.12 (1.07-4.20)

Supplementary Figures

Figure S1. Directed acyclic graph (DAG) illustrating the assumed causal pathways involving exposure to $PM_{2.5}$ and mortality in the presence of specific sets of other risk factors. The figure distinguishes between area-level and individual-level sets of factors, and it includes both observed (definite) and unobserved (transparent) variables. The graph suggests that all the observed sets of factors (area-level deprivation, urban/rural & greenness, individual-level socio-economic variables, and lifestyle) should be adjusted for in order to avoid potential confounding.

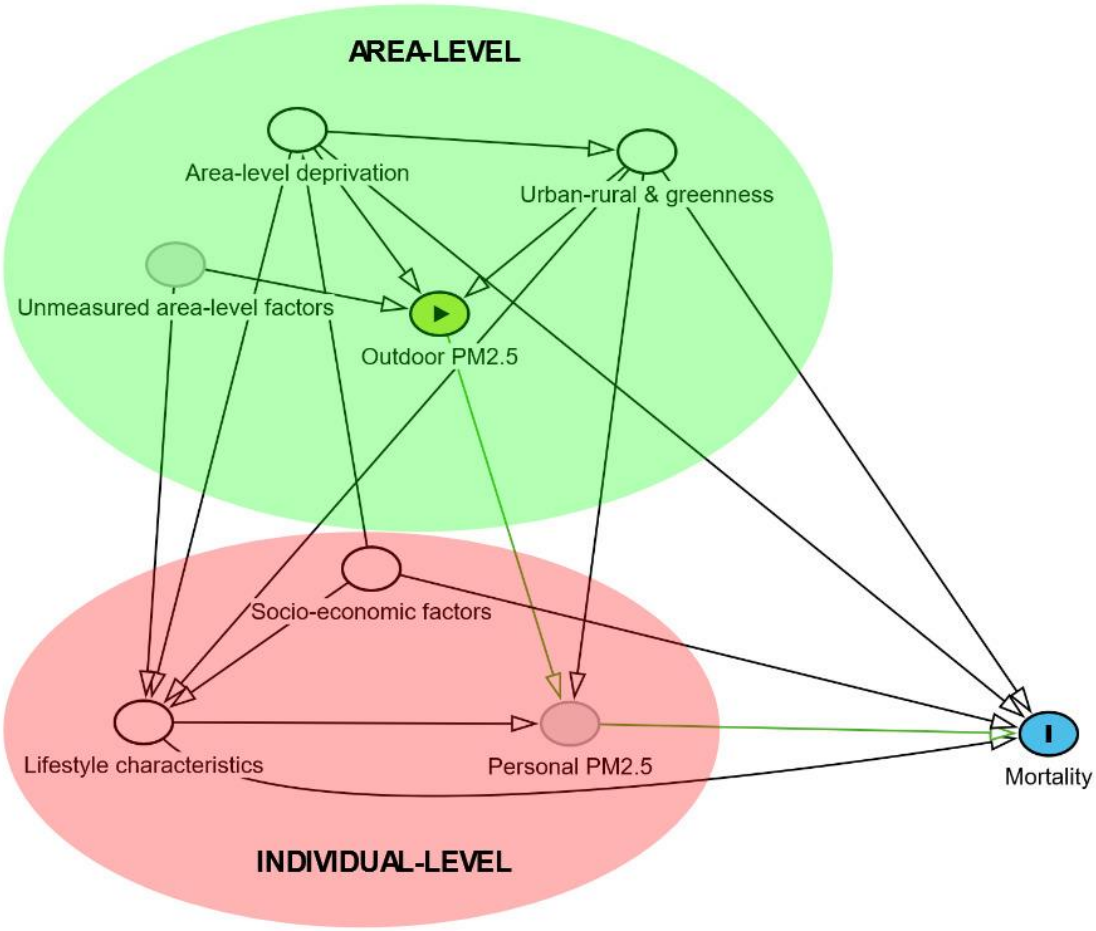
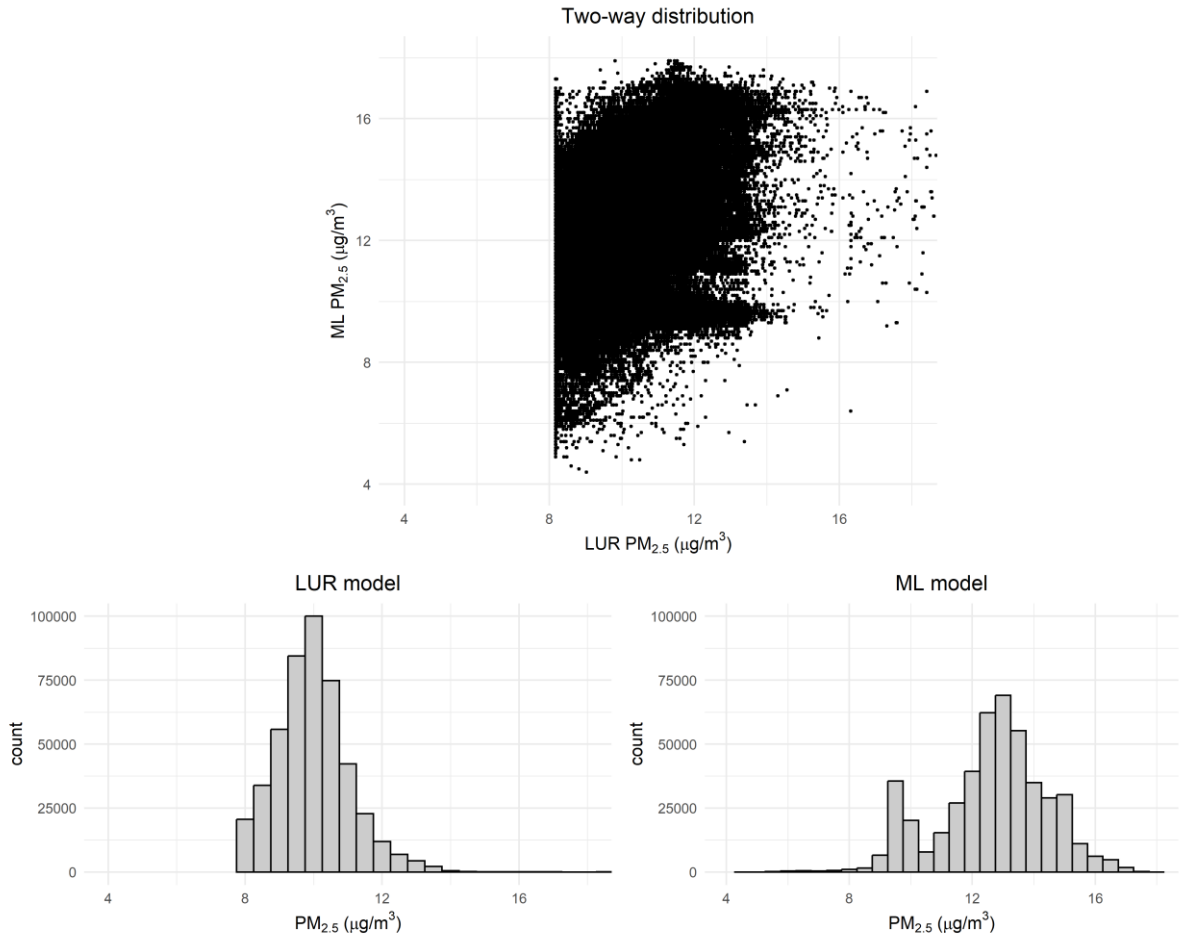


Figure S2. Comparison of distributions of the annual average exposure to $PM_{2.5}$ in 2010 across participants of the UKB cohort from the spatio-temporal machine learning (ML) model and the Europe-wide land-use regression (LUR) model.



Chapter 11

Research article 3

Association between long-term exposure to low ambient $PM_{2.5}$ and cardiovascular hospital admissions: a UK Biobank study

London School of Hygiene & Tropical Medicine
Keppel Street, London WC1E 7HT
T: +44 (0)20 7299 4646
F: +44 (0)20 7299 4656
www.lshtm.ac.uk

国立大学法人
長崎大学
NAGASAKI UNIVERSITY

RESEARCH PAPER COVER SHEET

Please note that a cover sheet must be completed for each research paper included within a thesis.

SECTION A – Student Details

Nagasaki Student No	59720005	**Title**	Association between long-term exposure to low ambient PM2.5 and cardiovascular hospital admissions: a UK Biobank study
LSHTM Student ID No	2004062		
First Name(s)	Jacopo		
Surname/Family Name	Vanoli		
Thesis Title	Assessment of the effect of air pollution on the UK Biobank cohort		
Nagasaki Supervisor(s)	Lina Madaniyazi		
LSHTM Supervisor(s)	Antonio Gasparrini		

If the Research Paper has previously been published please complete Section B, if not please move to Section C.

SECTION B – Paper already published

Where was the work published?	Environment International		
When was the work published?	2024		
If the work was published prior to registration for your research degree, give a brief rationale for its inclusion			
Have you retained the copyright for the work?*	**Yes**	Was the work subject to academic peer review?	**Yes**

*If yes, please attach evidence of retention. If no, or if the work is being included in its published format, please attach evidence of permission from the copyright holder (publisher or other author) to include this work.

SECTION C – Prepared for publication, but not yet published

Where is the work intended to be published?	

Please list the paper's authors in the intended authorship order:	
Stage of publication	Choose an item.

SECTION D – Multi-authored work

For multi-authored work, give full details of your role in the research included in the paper and in the preparation of the paper. (Attach a further sheet if necessary)	**I took the lead in carrying out the analysis, structuring and writing up the paper**

SECTION E – Names and affiliations of co-author(s)

Please list all the co-authors' names and their affiliations.

Jacopo Vanoli[1,2], Arturo de la Cruz[2], Francesco Sera[3] , Massimo Stafoggia[4], Pierre Masselot[2,9], Malcolm Mistry[2,5], Sanjay Rajagopalan[6], Jennifer Quint[7], Chris Fook Sheng Ng[8], Lina Madaniyazi[1], Antonio Gasparrini[2,9]

[1]School of Tropical Medicine and Global Health, Nagasaki University, Nagasaki, Japan, [2]Department of Public Health Environments and Society, London School of Hygiene and Tropical Medicine, London, UK, [3]Department of Statistics, Computer Science, Applications, University of Florence, Florence, Italy, [4]Department of Epidemiology, Lazio Region Health Service, Rome, Italy, [5]Department of Economics, Ca' Foscari University of Venice, Venice, Italy, [6]Harrington Heart and Vascular Institute, University Hospitals Cleveland Medical center, Cleveland, OH, 44106, USA, [7]National Heart and Lung Institute, Imperial College London, London, UK, [8]Department of Global Health Policy, Graduate School of Medicine, The University of Tokyo, Tokyo, Japan, [9]Centre for Statistical Methodology, London School of Hygiene & Tropical Medicine, London UK

SECTION F

I confirm that all co-authors have agreed that the above paper will be included in my PhD thesis.

Student Signature	[redacted]
Date	9/5/2024

LSHTM Supervisor Signature	[redacted]
Date	05/09/2024

Nagasaki University Supervisor Signature	[redacted]
Date	05/09/2024

Please submit this form to: tmghadmin@ml.nagasaki-u.ac.jp

Environment International 192 (2024) 109011

Contents lists available at ScienceDirect

Environment International

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

Full length article

Association between long-term exposure to low ambient $PM_{2.5}$ and cardiovascular hospital admissions: A UK Biobank study

Jacopo Vanoli [a,b,*], Jennifer K. Quint [c], Sanjay Rajagopalan [d], Massimo Stafoggia [e], Sadeer Al-Kindi [f], Malcolm N. Mistry [b,g], Pierre Masselot [b], Arturo de la Cruz Libardi [b], Chris Fook Sheng Ng [h], Lina Madaniyazi [a], Antonio Gasparrini [b]

[a] School of Tropical Medicine and Global Health, Nagasaki University, Nagasaki, Japan
[b] Environment & Health Modelling (EHM) Lab, Department of Public Health Environments and Society, London School of Hygiene & Tropical Medicine, London, UK
[c] School of Public Health, Imperial College London, London, UK
[d] Harrington Heart and Vascular Institute, University Hospitals Cleveland Medical Center, Cleveland, OH 44106, United States
[e] Department of Epidemiology, Lazio Region Health Service ASL ROMA 1, Rome, Italy
[f] Center for Health and Nature, Houston Methodist, Houston, TX, United States
[g] Department of Economics, Ca' Foscari University of Venice, Venice, Italy
[h] Department of Global Health Policy, Graduate School of Medicine, The University of Tokyo, Tokyo, Japan

ARTICLE INFO

Keywords:
Air pollution
Cardiovascular disease
Hospital admissions
Long-term

ABSTRACT

Introduction: A causal link between air pollution exposure and cardiovascular events has been suggested. However fewer studies have investigated the shape of the associations at low levels of air pollution and identified the most important temporal window of exposure. Here we assessed long-term associations between particulate matter < 2.5 μm ($PM_{2.5}$) at low concentrations and multiple cardiovascular endpoints using the UK Biobank cohort.

Methods: Using data on adults (aged > 40) from the UK Biobank cohort, we investigated the associations between 1-year, 3-year and 5-year time-varying averages of $PM_{2.5}$ and incidence of major adverse cardiovascular events (MACE), myocardial infarction (MI), heart failure, atrial fibrillation and flutter and cardiac arrest. We also investigated outcome subtypes for MI and stroke. Events were defined as hospital inpatient admissions. We fitted Cox proportional hazard regression models applying extensive control for confounding at both individual and area level. Finally, we assessed the shape of the exposure–response functions to assess effects at low levels of exposure.

Results: We analysed data from 377,736 study participants after exclusion of prevalent subjects. The average follow-up (2006–2021) was 12.9 years. We detected 19,353 cases of MACE, 6,562 of acute MI, 6,278 of heart failure, 1,258 for atrial fibrillation and flutter, and 16,327 for cardiac arrest. Using a 5-year exposure window, we detected positive associations (for 5 $\mu g/m^3$ increase in $PM_{2.5}$) for 5-point MACE of [1.12 (95 %CI: 1.00–1.26)], heart failure [1.22 (1.00–1.50)] and cardiac arrest [1.16 (1.03–1.31)]. We did not find any association with acute MI, while non-ST-elevation MI was associated with the 1-year exposure window [1.52 (1.12–2.07)]. The assessment of the shape of the exposure–response relationships suggested that risk is approximately linear for most of the outcomes.

Conclusions: We found positive associations between long-term exposure to $PM_{2.5}$ and multiple cardiovascular outcomes for different exposure windows. The cardiovascular risk tends to rise even at exposure concentrations below 12–15 $\mu g/m^3$, indicating high risk below UK national and international thresholds.

* Corresponding author.
E-mail addresses: jacopo.vanoli@lshtm.ac.uk (J. Vanoli), j.quint@imperial.ac.uk (J.K. Quint), sxr647@case.edu (S. Rajagopalan), m.stafoggia@deplazio.it (M. Stafoggia), sal-kindi@houstonmethodist.org (S. Al-Kindi), Malcolm.Mistry@lshtm.ac.uk (M.N. Mistry), Pierre.Masselot@lshtm.ac.uk (P. Masselot), Arturo.de-la-Cruz-Libardi@lshtm.ac.uk (A. de la Cruz Libardi), chrisng@m.u-tokyo.ac.jp (C. Fook Sheng Ng), lina.madaniyazi@nagasaki-u.ac.jp (L. Madaniyazi), Antonio.Gasparrini@lshtm.ac.uk (A. Gasparrini).

https://doi.org/10.1016/j.envint.2024.109011
Received 28 May 2024; Received in revised form 22 August 2024; Accepted 12 September 2024
Available online 14 September 2024

1. Introduction

Historically, cardiovascular events have been among the most prominent contributors to the global burden of disease, causing 4.75 million deaths annually (Vaduganathan et al., 2022). Thanks to significant scientific progress in medical therapies, preventive measures, and increased public awareness over the years, high-income countries have experienced a decline in adverse clinical endpoints related to cardiovascular issues. However, globally, cardiovascular diseases continue to pose a substantial burden and remain a primary concern for national healthcare systems (Cheema et al., 2022).

Air pollution is a well-recognized risk factor for cardiovascular diseases and, among other air pollutants, $PM_{2.5}$ (particulate matter with an aerodynamic diameter less than 2.5 μm) is known to be the most detrimental.

$PM_{2.5}$ has been suggested to be causally related to cardiovascular disease and modulates its effects through a multitude of mechanisms, including progression of atherosclerosis and promotion of vulnerable plaque (Sanjay et al., 2018). Both acute plaque instability and chronic progression of plaque may ultimately result in the presentation of acute myocardial infarction and stroke. Therefore, considerable attention has been devoted to understanding the timelines of exposure to air pollution and resultant cardiovascular events (Al-Kindi et al., 2020; Crouse et al., 2020). There is also an emerging body of evidence linking antecedent exposure to air pollution with heart failure hospitalization and arrhythmias, including both atrial fibrillation and ventricular arrhythmias (de Bont et al., 2022). However, there is a paucity of studies on the relevant temporal windows of exposure (Crouse et al., 2020).

Further, studies that have evaluated a variety of composite outcomes are rare (Osborne et al., 2023), and the presence of non-linear effects, especially at lower levels of exposure, has been investigated in detail only in a few investigations with only one in Europe (Wolf et al., 2021; Danesh Yazdi et al., 2019; Bai et al., 2019).

In this study, we made use of the wealth of individual-level data present in UK Biobank (UKB) cohort, paired it with temporally resolved ambient $PM_{2.5}$ exposure data, to assess risks of hospitalizations for several cardiovascular outcomes. We aimed to explore long-term associations at low levels of $PM_{2.5}$ using time-varying averages at multiple temporal windows of exposure. We also assessed non-linear effects by varying the shape of the exposure–response function and by restricting the analysis to subjects with exposures below predefined thresholds.

2. Methods

2.1. Population (UK Biobank cohort)

The British prospective cohort study, UK Biobank (UKB), enrolled approximately half a million individuals aged 40 to 69 years between 2006 and 2010. As a first cohort assessment, the participants underwent an in-person visit in one of the 22 assessment centres located across Great Britain (England, Scotland, and Wales). The visit included multiple questionnaires regarding lifestyles and personal characteristics. Anthropometric measures and biological samples were also collected. Participants were followed up through periodical linkage with administrative health databases, including mortality and cancer national registries as well as primary and secondary care visits. The cohort profile has been described in detail in previous publications (Sudlow et al., 2015; Fry et al., 2017). Specific details regarding the UKB database can be found on the showcase website (https://biobank.ndph.ox.ac.uk/showcase/).

2.2. Study design

This study followed a time-to-event design. In this analysis, we excluded subjects with cardiovascular hospital admission prior to enrolment. Subsequently, we excluded subjects who, at enrolment time, self-reported prior cardiovascular diseases and/or hypertension medication. Participants were censored at the date of event occurrence, date of death, loss to follow-up, or the administrative end of follow-up (set here to 31/12/2021), whichever came first.

2.3. Exposure assessment and linkage

We assigned exposure at individual level combining $PM_{2.5}$ predictions and residential history data. The original $PM_{2.5}$ data were represented by daily levels predicted on a 1-km grid across the UK in the period 2003–2021 using a hybrid spatio-temporal machine learning (ML) model. The model used an ensemble of ML algorithms trained using ground monitor series and a series of spatial and spatio-temporal predictors, including outputs from emission-dispersion models, remote sensing satellite data, as well as land-use and traffic variables, among others (de la Cruz et al., 2024). The model performance, assessed using cross-validation, provided an overall coefficient of determination (R^2) of 0.80 at daily scale. The residential data were available in the UKB database, including periods and geocoded locations with 100 m rounding. The data were validated internally, and the mobility history continuously updated through general practitioner registration or direct reporting by the participants.

The linkage process was performed in two steps. First, we constructed daily exposure series for each residential location by interpolating gridded exposure values using bilinear method. This approach consist of a linear interpolation over a two-dimensional grid and allowed preserving the exposure information while masking the original residential data, thus preventing back-tracing of the individual locations. Second, we composed the subject-specific exposure series by matching the daily series for corresponding residential periods. The process has been described in detail in a previous publication (Vanoli et al., 2024).

2.4. Hospital admissions outcomes

At the time of enrolment, people consented for access to a variety of personal information, including linked electronic health records and residential address locations. The UKB provides access to summary datasets including first inpatient hospital visits and operation codes. For each outcome, ICD-10 code and date of first primary or secondary diagnosis are made available. In this analysis, we used codes for the following outcome diagnoses: heart failure (I50.x, where ".x" defines all code subtypes), atrial fibrillation and flutter (I46, I46.0, I46.1, I46.9), cardiac arrest (I48, I48.0, I48.1, I48.2, I48.3,I 48.4), acute myocardial infarction (I21.x, I23.x), ST-elevation myocardial infarction (STEMI, I21.0–3), non-ST-elevation myocardial infarction (NSTEMI, I21.4), intracerebral stroke (I61.x), ischaemic stroke (I63.x, I64.x), and subarachnoid stroke (I60.x). In addition, we created a composite major adverse cardiovascular event (5-point MACE) outcome, defined as the occurrence of either acute MI (I21.x, I23.x), stroke (I60.x,I61.x,I63.x, I.64.x), unstable angina (I20.0) and heart failure (I50.x) and death due to cardiovascular disease (I00-I99). Details on the outcome diagnoses' definitions can be found on the UKB website (https://biobank.ndph.ox.ac.uk/ukb/label.cgi?id = 2002).

2.5. Statistical analysis

We constructed separate cohorts to analyze each outcome based on an extended Cox proportional hazard model for time-varying exposures where the follow-up of each subject was split by calendar year. Therefore, we performed the analysis based on an extended Cox proportional hazard model with time-varying exposure (Andersen and Gill, 1982). We defined the model using calendar years as timescale and we stratified by assessment centre, sex, and year of birth, thus ensuring control for differential risks by age. The extended survival data was linked with annual exposure averages assigned over a lag window of five years, from lag 0 (the year of the event) until lag 4 (i.e. fourth year before the year of

the event), consistently with previous studies (Crouse et al., 2015). Subjects with incomplete exposure history were excluded from the analysis.

In the main analysis, the exposure term was defined as the time-dependent average across the lag periods (in years) and we assumed a linear exposure–response relationship. Specifically, we investigated the associations for lag 0 (1-year average), lag 0–2 (3-year average) and lag 0–4 (5-year average) in separate models, due to their potentially high correlation. Additionally, we evaluated the shape of the association between time-varying $PM_{2.5}$ for lag 0–4 and each outcome by estimating a non-linear response function using penalized splines, with the optimal degrees of freedom selected using the Akaike Information Criterion (AIC).

We a priori specified two confounder models: *Model 1* included the matching variables used for stratification (assessment centre, sex, and year of birth) and individual covariates determined at recruitment: ethnic background education level, household income, employment status, smoking status, packs of cigarettes per year, average alcohol intake per week, waist-to-hip ratio, physical activity (measured using the International Physical Activity Questionnaire (IPAQ) scale) and living alone (a proxy for marital status). In *Model 2*, we added control for area-level covariates, including the Townsend Deprivation Index measured in 2010, urban–rural classification (urban, town or fringe and village), and greenness percentage around the baseline residential address (at 1000 m buffer based on the UKB internal definition (Generalised Land Use Database Statistics for England, 2005; Morton et al., 2011)).

Estimates of the associations were reported as hazard ratios (HRs) for each hospitalization outcome per 5 $\mu g/m^3$ increments in $PM_{2.5}$, with 95 % confidence intervals (CI). Missing values in the baseline covariates were imputed using multiple imputation by chained equation (MICE), producing five imputed datasets, with estimates combined using Rubin's rule (Barnard and Rubin, 1999).

2.6. Additional analyses

We performed a sensitivity analysis including only person-years assigned to exposure levels below 10 (WHO 2005 limits) or 12 $\mu g/m^3$, as done previously (Wolf et al., 2021). We did not investigate associations at exposure levels below the WHO 2021 annual limits (5 $\mu g/m^3$) because of the paucity of data in that exposure range. To evaluate sensitivity in the associations due to different MACE definitions, we conducted sensitivity analyses using 4- and 3-point MACE. Those were defined as the 5-point MACE outcome sequentially excluding diagnosis for heart failure (4-point MACE) and unstable angina (3-point MACE), respectively. To assess potential changes in the association attributable to the COVID-19 period we conducted an analysis with follow-up up to 31/12/2019. We conducted an additional analysis subsetting the cases only to the participants who had the diagnosis in the primary position. We also reported results by IQR increase (3.7 $\mu g/m^3$) to make our results more in line with the exposure distribution. Finally, we performed an analysis including a washout period in order to account for potential healthy-volunteer and other selection biases, as recommended in a recent publication (Chen et al., 2024).

Data cleaning and statistical analyses were conducted using in R Statistical Software (version 4.2.1) and the following packages were used: data.table, survival, mice, parallel, ggplot2 and gridExtra.

3. Results

3.1. Descriptives

The original dataset included 502,381 individuals. We excluded subjects with cardiovascular inpatient hospital admission prior to enrolment (n = 83,249), with self-reported history cardiovascular disease (n = 8,491) and hypertension medication (n = 32,664).Finally, 241 participants were excluded due to (partially) missing exposure histories, providing a final cohort of 377,736 individuals (Fig. 1). The participants were followed-up for an average of 12.9 years, with a total of 4,877,026 person-years. During the follow-up, among all the participants, 6,278 had inpatient hospital visit due to heart failure, 1,258 of atrial fibrillation and flutter, 16,327 of cardiac arrest, 6,562 of acute myocardial infarction, 2,710 of MI STEMI 2,426 of MI NSTEMI, 928 of intracerebral stroke, 4,526 of ischaemic stroke, 664 of subarachnoid stroke. For the composite outcomes, 5-point MACE status was reported for 19,353 participants.

The cohort had slightly more females than males (See Table 1), with an average age of 55 at baseline, and most of the cohort was of white ethnicity. More than 70 % of the subjects had at least received a diploma and 60 % were employed at the time of recruitment. About 11 % of participants were smokers, approximately half of the rate in the general UK population in 2011(General Lifestyle Survey:, 2011). Most subjects (84 %) lived in urban settings. In the proximity of the residential address, the average greenspace percentage and the average Townsend deprivation index were 45 and −1.39, respectively (See Table 1). These values reflect relatively wealthy residential surroundings.

Figure S1 showed the box-and-whiskers plot of the distribution of annual averages of $PM_{2.5}$ across the years from 2007 until 2021. The plot indicated that all UKB participants are permanently exposed to exposure values below the UK Air Quality Objectives (AQO) and EU Air Quality Directives (AQD) for 2015 and 2020 of 25 $\mu g/m^3$. After a slight increase in 2011, the distribution of $PM_{2.5}$ had generally declined over time. Since 2015, the majority of the cohort has been exposed to levels below the the 2005 WHO Air Quality Guidelines (AQG) limit of 10 $\mu g/m^3$. Seldom annual exposure levels were below the new WHO AQG 2020 limits of 5 $\mu g/m^3$. The correlation matrix (Table S7) between the exposure windows showed high to very high correlation among the exposure windows.

3.2. Associations between CVD and $PM_{2.5}$ exposure

In Table 2, we showed the linear associations, reported as hazard ratios (HRs) for a 5 $\mu g/m^3$ increase, between $PM_{2.5}$ exposure with different lag windows and each cardiovascular outcome. In the fully-adjusted model (Model 2), the exposure was significantly associated with elevated risk for diagnosis of heart failure, intracerebral stroke, cardiac arrest and MI NSTEMI. For example, using an exposure window with lag 0–4, heart failure displayed an HR of 1.22 (95 %CI: 1.00–1.50) and intracerebral stroke of 1.94 (1.15–3.29). Associations were also positive for 5-point MACE (1.12 (1.00–1.26) at lag 0–4) but we found the strongest effects for shorter exposure windows (lag 0–2, 1.15 (1.03–1.28)). We did not find any evidence of linear associations with ischaemic stroke, subarachnoid stroke, acute MI, and atrial fibrillation and flutter. In general, associations for several outcomes were positive but did not reach statistical significance at the 5 % level, probably due to limited statistical power.

The comparison between Model 1 and 2 indicates that the inclusion of area-level confounders led to an attenuation of the estimates, except for MI NSTEMI, for which the associations increased.

Overall, we did not find important differences in the associations across different exposure windows, with some exceptions. For instance, the increased risk for MI NSTEMI was significant only when we considered the exposure of the last year (lag0). In contrast, for cardiac arrest, the lag0 window showed a weak relationship, while the other windows had stronger associations.

Linear associations in Table 3 and S3 compared the main associations with those estimated for subsets of person-years exposure to low concentrations (<=10 and <=12). The results mostly showed the strongest effects below a concentration of 10 $\mu g/m^3$, but the associations below 12 are more difficult to interpret and show unclear patterns across the outcomes. This is likely due to the uneven distribution of $PM_{2.5}$ (figure S1) across calendar years, which showed concentrations higher than 12

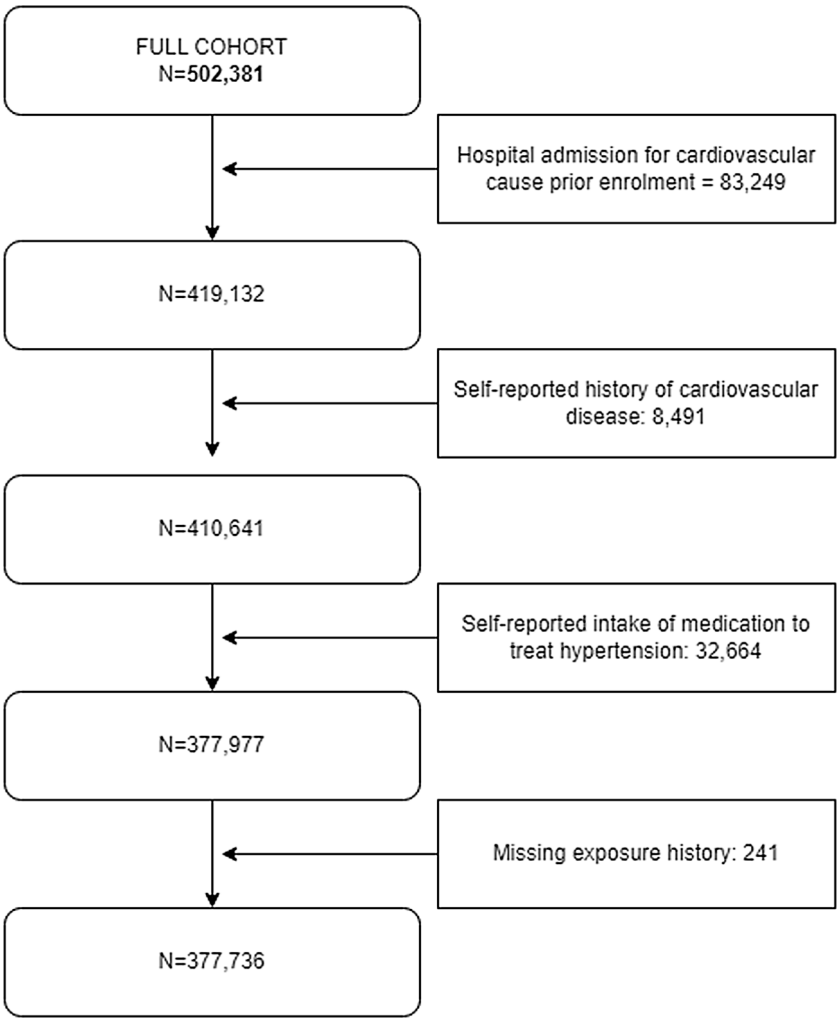


Fig. 1. Flow diagram representing the selection of the sample of the UK Biobank.

in the first years of exposure (up to 2015) when likely few events had occurred and therefore the corresponding estimates are more uncertain.

Model allowing non-linear associations (Fig. 2) showed limited evidence for deviations from linear exposure–response function associations. For several outcomes, including 5-point MACE, MI STEMI, intracerebral stroke, heart failure, atrial fibrillation and flutter, and cardiac arrest, the penalized spline indicated linear associations, confirmed by the non-significant Wald test for non-linearity. In contrast, acute MI showed a non-linear effect, with the curve increasing up to 12–13 $\mu g/m^3$ before declining. There was no evidence for a threshold in the association. No significant effects were observed for ischemic and subarachnoid stroke.

3.3. Additional analyses

Analysis of codes in only primary position (Table S1) showed similar associations with both primary and secondary code analysis (Table 2). Associations estimated with follow-up up to 31/12/2019 (Table S2) are mostly consistent with the main results (Table 2). In general, strong differences in the HR and confidence intervals between main and sensitivity analyses can be attributed to reduced sample sizes, considering that some outcomes only have a few hundred events in total and subsetting can lead to instabilities in the associations. Among MACE outcomes, 5-point MACE exhibited the highest and more precise associations compared to 3- and 4-point MACE (Table S4), likely due to increased statistical power.

For most of the outcomes, the exclusion of the wash-out period (Table S6) lead to stronger positive associations compared to the main linear analysis. Contrarily, subarachnoid and ischaemic stroke still displayed null effects.

4. Discussion

In this 15-year UK-based study, we used state-of-the-art epidemiological methods to assess the association between chronic exposure to time-varying $PM_{2.5}$ at different yearly time windows and risk of hospitalizations for MACE and other severe clinical cardiovascular endpoints. We observed positive linear associations between $PM_{2.5}$ across multiple exposure windows and several outcomes, including 5-point MACE, heart failure, intracerebral stroke and cardiac arrest. On the other hand, we found significant non-linear associations with overall acute MI and MI NSTEMI.

This study aimed at addressing research recommendations issued by the COMEAP 2019 report on air pollution and cardiovascular diseases (Kelly, 2019). First, the report highlighted the need for the use of more refined exposure estimates: in this analysis we applied highly resolved predictions from a state-of-the-art exposure model for the first time in the UK. Second, we investigated both major and subtypes of outcomes, non-linear effects and different exposure windows, with the purpose to shed more light on the mechanistic effects of long-term exposure on cardiovascular diseases, a question that was also part of the research recommendations.

Table 1
Descriptive statistics for continuous (mean with 5th-95th percentile range and missing) and categorical (counts/percentage) baseline characteristics and number of outcomes' events in the UKB cohort.

sex	Female	211,843 (56.1 %)
	Male	165,893 (43.9 %)
	Missing (%)	0 (0.0 %)
ethnicity	White	356,597 (94.4 %)
	Other	19,767 (5.2 %)
	Missing (%)	1,372 (0.4 %)
employment status	Employed	235,367 (62.3 %)
	Retired	108,985 (28.9 %)
	Other	29,730 (7.9 %)
	Missing (%)	3,654 (1.0 %)
educational level	Low	54,078 (14.3 %)
	Professional Qualification	41,721 (11.0 %)
	Highschool diploma	144,801 (38.3 %)
	College/University degree	130,347 (34.5 %)
	Missing (%)	6,789 (1.8 %)
household income	Less than 18,000	64,043 (17.0 %)
	18,000 to 30,999	79,325 (21.0 %)
	31,000 to 51,999	88,037 (23.3 %)
	Greater than 100,000	19,863 (5.3 %)
	Missing (%)	54,165 (14.3 %)
physical activity (IPAQ score)	low	55,596 (14.7 %)
	moderate	125,054 (33.1 %)
	high	125,601 (33.3 %)
	Missing (%)	71,485 (18.9 %)
alcohol intake	Never	27,563 (7.3 %)
	Special occasions only	41,122 (10.9 %)
	One to three times a month	42,435 (11.2 %)
	Once or twice a week	98,996 (26.2 %)
	Three or four times a week	89,992 (23.8 %)
	Daily or almost daily	77,229 (20.4 %)
	Missing (%)	399 (0.1 %)
smoking status	Never	212,625 (56.3 %)
	Previous	123,450 (32.7 %)
	Current	40,332 (10.7 %)
	Missing (%)	1,329 (0.4 %)
living alone	No	308,249 (81.6 %)
	Yes	67,837 (18.0 %)
	Missing (%)	1,650 (0.4 %)
urban/rural classification	Urban	317,266 (84.0 %)
	Town/fringe	28,263 (7.5 %)
	Village/Rural	28,204 (7.5 %)
	Missing (%)	4,003 (1.1 %)
age at baseline	Years	55.46 (42.00 to 68.00)
	Missing (%)	0 (0.0 %)
waist-to-hip ratio		0.86 (0.72 to 1.01)
	Missing (%)	1,210 (0.3 %)
Smoking intensity	packs-year	7.27 (0.00 to 37.50)
	Missing (%)	57,592 (15.2 %)
Townsend deprivation index (2010		−1.39 (−5.06 to 4.77)
	Missing (%)	467 (0.1 %)
Greenspace	percentage	45.09 (15.40 to 87.38)
	Missing (%)	45,206 (12.0 %)
Number of events		
3-point MACE		14,087
4-point MACE		15,186
5-point MACE		19,353
acute MI		6,562
MI STEMI		2,710
MI NSTEMI		2,426
intracerebral stroke		928
ischaemic stroke		4,526
subarachnoid stroke		664
heart failure		6,278
atrial fibrillation and flutter		1,258
cardiac arrest		16,327

Table 2
Hazard ratios (HRs, with 95 % confidence intervals) of cardiovascular outcomes associated with an increase of 5 μg/m^3 in $PM_{2.5}$ in the UKB cohort, for combinations of length of exposure windows (lag0, lag02, lag04) and confounding control.

Outcome	exposure window	Model 1	Model 2
5-point MACE	lag0	1.17 (1.08–1.28)	1.13 (1.02–1.24)
	lag02	1.20 (1.09–1.32)	1.15 (1.03–1.28)
	lag04	1.18 (1.07–1.31)	1.12 (1.00–1.26)
Myocardial Infarction (MI)			
Acute	lag0	0.98 (0.84–1.14)	1.07 (0.90–1.27)
	lag02	0.95 (0.81–1.12)	1.06 (0.87–1.28)
	lag04	0.95 (0.81–1.13)	1.06 (0.87–1.29)
STEMI	lag0	0.89 (0.71–1.12)	0.98 (0.75–1.28)
	lag02	0.91 (0.71–1.17)	1.03 (0.77–1.38)
	lag04	0.94 (0.73–1.22)	1.09 (0.80–1.49)
NSTEMI	lag0	1.17 (0.90–1.51)	1.52 (1.12–2.07)
	lag02	1.04 (0.80–1.36)	1.35 (0.97–1.88)
	lag04	1.03 (0.78–1.35)	1.32 (0.95–1.84)
Cerebrovascular disease and stroke			
Intracerebral stroke	lag0	1.81 (1.20–2.72)	1.74 (1.09–2.78)
	lag02	1.93 (1.26–2.97)	1.91 (1.15–3.17)
	lag04	1.96 (1.26–3.05)	1.94 (1.15–3.29)
Ischaemic stroke	lag0	1.14 (0.95–1.37)	1.07 (0.87–1.32)
	lag02	1.16 (0.96–1.41)	1.08 (0.86–1.36)
	lag04	1.12 (0.91–1.36)	1.01 (0.80–1.28)
Subarachnoid stroke	lag0	1.02 (0.64–1.61)	0.94 (0.56–1.56)
	lag02	1.08 (0.66–1.76)	0.98 (0.56–1.74)
	lag04	1.12 (0.67–1.88)	1.03 (0.56–1.89)
Other outcomes			
Heart failure	lag0	1.35 (1.15–1.58)	1.19 (1.00–1.42)
	lag02	1.40 (1.18–1.65)	1.21 (1.00–1.48)
	lag04	1.41 (1.19–1.68)	1.22 (1.00–1.50)
Atrial fibrillation and flutter	lag0	1.43 (1.01–2.01)	1.39 (0.94–2.05)
	lag02	1.43 (0.99–2.07)	1.38 (0.90–2.12)
	lag04	1.34 (0.92–1.96)	1.26 (0.81–1.95)
Cardiac arrest	lag0	1.13 (1.03–1.24)	1.09 (0.98–1.22)
	lag02	1.17 (1.06–1.30)	1.15 (1.02–1.29)
	lag04	1.19 (1.07–1.31)	1.16 (1.03–1.31)

Table 3
Hazard ratios (HRs, with 95 % confidence intervals) of cardiovascular outcomes in exposure subset analysis in the UKB cohort. Exposure is defined as lag04 (5-years time-dependent average). Results are for an increase of 5 μg/m^3 in $PM_{2.5}$ using fully adjusted models (Model 2).

Outcome	Exposure subset		
	<= 10	<= 12	Full analysis
5-point MACE	1.05 (0.84–1.32)	1.16 (1.01–1.34)	1.12 (1.00–1.26)
Myocardial Infarction (MI)			
Acute	1.27 (0.90–1.81)	0.78 (0.53–1.17)	1.06 (0.87–1.29)
STEMI	1.20 (0.94–1.53)	1.07 (0.81–1.43)	1.09 (0.80–1.49)
NSTEMI	1.19 (0.68–2.07)	0.75 (0.25–2.22)	1.32 (0.95–1.84)
Cerebrovascular disease and stroke			
Intracerebral stroke	1.70 (1.01–2.86)	1.10 (0.89–1.36)	1.94 (1.15–3.29)
Ischaemic stroke	1.45 (1.00–2.10)	0.98 (0.85–1.14)	1.01 (0.80–1.28)
Subarachnoid stroke	0.98 (0.74–1.29)	1.33 (0.95–1.87)	1.03 (0.56–1.89)
Other outcomes			
Heart failure	1.51 (0.58–3.93)	1.13 (0.95–1.35)	1.22 (1.00–1.50)
Atrial fibrillation and flutter	0.58 (0.27–1.22)	1.22 (0.98–1.52)	1.26 (0.81–1.95)
Cardiac arrest	0.95 (0.55–1.64)	1.08 (0.93–1.25)	1.16 (1.03–1.31)

4.1. MACE

This is one of the first studies to investigate the effect of long-term air pollution on MACE, and we found positive associations. Across all the time windows, UK resident adults living in areas with a 5 μg/m^3 higher exposure experienced a 12 % to 15 % elevated risk of MACE-related hospitalizations compared to those in less exposed areas. The direction of the effect is in line with two studies on US veterans where a 9 % (by a 5 μg/m^3 increase in $PM_{2.5}$) increased risk was observed in individuals with prior coronary artery bypass grafting (Deo et al., 2024), and a 52 % increased risk in those with prior percutaneous coronary interventions (Motairek et al., 2023). In contrast, a Swedish (Carlsen et al., 2022) study showed no significant associations. It is important to consider that our studies differ in MACE composition, as our analysis included various clinical events, while, for example, the Swedish study focused on myocardial infarction and coronary interventions, making direct comparisons challenging. In our study, we attempted to include a simple and unambiguous MACE definition, therefore we decided to use the outcome as described in a recent literature review (Bosco et al., 2021). We hope that our choice may be of help to define outcomes in other medium-sized cohort studies that require composite definitions to conduct well-powered analyses.(Bosco et al., 2021).

4.1.1. Myocardial infarction

Our findings on acute myocardial infarction (MI) show positive but not statistically significant associations [lag04: 1.07 (0.90–1.27)], aligning with two large *meta*-analyses (Alexeeff et al., 2021; Zhu et al., 2021) and a Canadian study (Bai et al., 2019). At low pollution levels, recent studies also reported positive effects (Wolf et al., 2021). However, while a Europe-based cohort reported similar associations with our study [1.02 (0.95–1.10)] (Wolf et al., 2021), US studies found stronger associations in the HR range of 1.13–1.188(Danesh Yazdi et al., 2019; Alexeeff et al., 2023).

This is one of the first long-term studies to examine associations with two types of myocardial infarction: STEMI and NSTEMI. Both outcomes were associated with the exposure, but we found no statistical significance with STEMI while there was a positive significant association with NSTEMI. The strong effects observed for the latter suggest that this subtype of MI may play a significant role in the overall association with acute MI in this cohort. A reason for this could be that NSTEMI is defined as partial blockages in a coronary artery, leading to less severe but more widespread myocardial ischemia. This condition makes it more vulnerable to factors that can worsen systemic inflammation and endothelial dysfunction, such as chronic exposure to air pollution. On the other hand, STEMI usually results from a complete and abrupt blockage of a coronary artery, causing a more acute event, which may be more associated with short-term air pollution (Newby et al., 2015; Kuźma et al., 2024). Even though the literature is not completely consistent in their distinction (STEMI and NSTEMI, 2024), our post-hoc hypothesis has been previously supported by the data as STEMI, but not NSTEMI, has been associated with short-term exposure (Gardner et al., 2014; Pope et al., 2015). There are currently no long-term studies considering their distinction. On the other hand, the use of a relatively young cohort (under 65–70 years) might also partially explain the increased risk of NSTEMI due to air pollution. While STEMI has traditionally been associated with younger individuals compared to NSTEMI, there has been a recent trend reversal in the UK, where since 2016, younger people (under 65) are more frequently admitted with NSTEMI than older individuals (MINAP and NAPCI, 2023). Additionally, NSTEMI cases are often underrepresented in hospitals due to their lower severity (How the NHS Cares for Patients with Heart Attack, 2015), suggesting that the actual numbers may be higher than reported for mid-aged adults. In this context, air pollution exposure may disproportionately increase the risk for younger individuals as they likely spend more time outdoor but have less severe outcomes. Therefore, while older adults are more vulnerable to air pollution effects, mid-age adults may be more at risk of slowly developing atherosclerotic plaques that lead to worse outcomes later in life, such as a case of STEMI. Supporting this, a recent large Polish study investigating mid-term (30 days) effects also found an increased risk of NSTEMI in younger individuals (under 65)(Kuźma et al., 2024).

Finally, a similarity between our investigation and the Polish study is that both analyses identified significant effects of mid-term air pollution exposure (ranging from one month to one year) on NSTEMI, indicating a potential impact of air pollution over this duration. However, they do not reveal differences substantial enough to support any significant post-hoc pathophysiological hypotheses. Nonetheless, these findings suggest that it may be valuable to incorporate sensitivity analyses with varying exposure windows in long-term studies, similar to those used in short-term research.

4.1.2. Stroke

A large body of literature has covered the air pollution-stroke events relationship. In comparison to two *meta*-analyses that focused on general cerebrovascular disease (Alexeeff et al., 2021; Niu et al., 2021), our effects were higher for intracerebral stroke. Our analysis also indicates increased associations when compared with recent UK-(Cai et al., 2018); (Atkinson et al., 2013) and European-based (Wolf et al., 2021) studies. For example, Cai and colleagues (Cai et al., 2018), using the UKB cohort found null associations both for overall cerebrovascular diseases and stroke types (ischaemic and haemorrhagic). In this study we found heterogeneity of risk among stroke types. While several studies have assessed short-term associations, to our knowledge our study is among the first to investigate long-term effects on intracerebral stroke as a separate outcome (Verhoeven et al., 2021). This is relevant, as in the literature cerebrovascular events are often considered as a whole, although different stroke types are clinically considered as different diseases with separate etiologies (Verhoeven et al., 2021). There are physiological channels that connect exposure to air pollution might

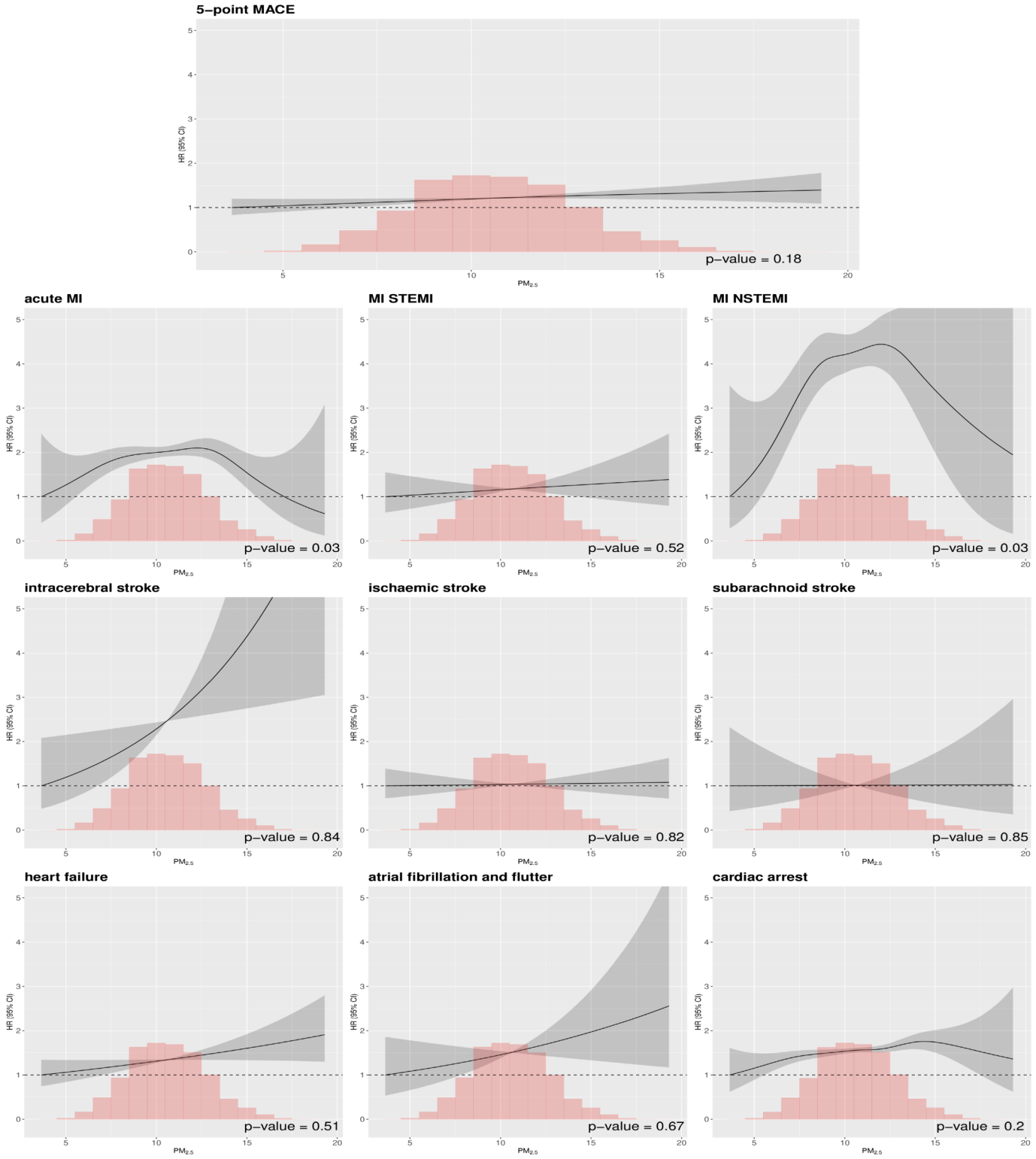


Fig. 2. Concentration-response functions of the associations between lag04 (5-year time-dependent) of $PM_{2.5}$ with cardiovascular events in the UK Biobank cohort. Models were fully adjusted (Model 2). The associations representing hazard ratios (HRs, with 95% confidence intervals) were estimated using penalized splines with degrees of freedom selected using the Akaike information criterion (AIC, solid line) with 95% confidence intervals (surrounding dashed lines).

affect the occurrence of intracerebral stroke but no other types of haemorrhagic stroke: first, through known mechanisms, that is inflammation, oxidative stress, and endothelial dysfunction, fine particles effects may be more pronounced on the small vessels within the brain compared to other larger arteries. However, a previous MRI study did not find effects of PM on markers of small vessels disease (Power et al., 2018) and therefore the hypothesis is weak. Second, chronic particulate exposure may indirectly affect the brain through the autonomic respiratory reflex arcs as well as uptake of particles which can induce marked neuro-inflammation (Verhoeven et al., 2021). Finally, some authors hypothesized that overproduction of amyloid protein related to cerebral amyloid angiopathy may be the cause of intracerebral stroke (Wilker

et al., 2018).

To conclude, the literature on the effects of air pollution on outcome subtypes is scattered and heterogeneous owing to the use of diverse study designs, exposure and outcome definitions and spatial differences in the particulate composition. More studies using state-of-the-art methodologies and harmonized datasets should be used to draw firmer hypotheses.

4.1.3. Heart failure, atrial fibrillation and cardiac arrest

We detected significant between long-term air pollution and arrhythmias related-outcomes [for lag04: 1.13 (1.02–1.24)]. This is in line with recent investigations on atrial fibrillation on Medicare data (Mahdieh Danesh et al., 2021) and a Canadian cohort (Shin et al., 2019). The latter also assessed the shape of the exposure–response relationship, showing no evidence of an effect of $PM_{2.5}$ below 6 $\mu g/m^3$, in contrast with our findings. A *meta*-analytical estimate on four older studies showed a null association [0.96 (0.82–1.13) for 5 $\mu g/m^3$ increase] (Pranata et al., 2020). Last, our associations for heart failure were positive, but on the boundaries of statistical significance[for lag04: 1.22 (1.00–1.50)], while a *meta*-analysis investigating the same outcome did not find effects. We found only one study (Shin et al., 2019) that examined hospital admissions related to atrial fibrillation and stroke, employing exposure windows similar to ours. No significant differences were observed for atrial fibrillation, whereas slight increases were noted for stroke for the 5-year window. However, our study suggests that the association may differ by exposure window.

4.1.4. Windows of exposure

The minor difference in estimates suggests different windows influence only moderately the health risks. For instance, exposure closer to the event (lag0) has a stronger impact on the risk for MACE and MI NSTEMI. On the other hand, longer exposure windows (lag 0–2 and lag 0–4) can be more important for stroke and cardiac arrest. Our comparison of varying window widths was similar only to two previous studies (Crouse et al., 2020; Lefler et al., 2019) that mainly focused on the risk of premature mortality with mixed evidence. In Lefler and colleagues (Lefler et al., 2019), results did not highlight any relevant window of exposure for cardiopulmonary deaths, while in the study by Crouse and colleagues (Crouse et al., 2020), longer exposure windows were consistently associated with increased risk of mortality both for ischaemic heart and cerebrovascular disease.

4.1.5. Shape of concentration–response function

The analysis of the concentration–response function suggest steep risks at concentrations even below 12–15 $\mu g/m^3$, with no evidence of a threshold at the lowest values. This result highlights that despite the recent decreases in the air pollution levels, air pollution carries adverse effects even at very low levels and therefore new mitigation strategies are needed to account for the public health burden that cannot be avoided by further lowering concentrations. This finding contributes to a growing body of literature emphasizing the significance of addressing air pollution concerns not only at elevated levels but also at lower exposure levels (Wolf et al., 2021; Chen et al., 2023; Di et al., 2017). Furthermore, for MI our non-linear estimates detected increased risks below 12 $\mu g/m^3$, agreeing with a previous study (Wolf et al., 2021). Contrarily, our results for stroke (intracerebral) suggest a linear relationship above 12 $\mu g/m^3$, while previous investigations found stronger associations, especially at low levels (Wolf et al., 2021; Shin et al., 2019). This may be due to the choice of outcomes' subtypes. Notably, the large majority of the previous literature focuses on US cohorts (Alexeeff et al., 2023; Di et al., 2017) while a few studies have investigated European cohorts (Wolf et al., 2021; Stafoggia et al., 2022).

4.2. Strengths and limitations

Our study carries several strengths. First, differently from the previous UK Biobank analyses of air pollution, this has been carried out using state-of-the-art exposure model with time-varying assignment, detailed confounders' information and statistical methodologies analogously to the most relevant air pollution studies in the literature to date. One of our study's main strengths lays in the utilization of time-dependent exposure summaries that enabled us to better define health risks compared to simpler exposure measures (Putter and van Houwelingen, 2017). This importantly distinguishes our approach from the majority of the UKB-based studies that solely incorporated fixed-time point exposures (Cai et al., 2018; Zhang et al., 2024; Luo et al., 2022; Parra et al., 2022) based on annual 2010 predictions of $PM_{2.5}$. Another key strength is the use of the sizable UKB cohort with a rich history of individual data. This allowed us to include important individual-level confounders in the models that are usually unaccounted for in air pollution studies, such as smoking and waist-to-hip ratio.

Furthermore, in this study, we incorporated a composite outcome in addition to specific endpoints. One of the benefits of using a composite endpoint instead of individual ones is the increased statistical power, resulting from the inclusion of a larger number of cases. This is evident in some of the results for specific cardiovascular disease (CVD) endpoints, where hazard ratios (HRs) are elevated but did not achieve statistical significance. Additionally, using a broader CVD definition rather than specific endpoints may reduce outcome measurement errors.

We used specific outcome types (e.g., ischaemic stroke) instead of general definitions (e.g., cerebrovascular disease). The diversity of health effects revealed in this study, particularly when examining subtypes of outcomes, underscores the importance of defining more detailed outcomes, instead of using a wide range of ICD codes. Specificity may be crucial in assisting clinicians both to pinpoint events strongly linked to exposure to air pollution and investigate the pathophysiological mechanisms of the diseases.

Finally, the long observation period in contrast to the majority of studies on cardiovascular outcomes could also be the reason for the differences in estimates between our research and existing literature.

Some limitations in our study should be highlighted. The primary limitation of the UKB cohort is the potential lack of representativeness of the UK population, possibly including to healthy-volunteer bias (Fry et al., 2017). To mitigate this issue, in a sensitivity analysis we defined a wash-out period (Chen et al., 2024), excluding person-years up to 2013. The results showed higher health effects compared to the main analysis for certain outcomes, suggesting that the original estimates might be conservative. Second, the use of administrative ICD codes to assess outcomes can be misleading, leading to diagnosis misclassification due to lack of clinical details regarding the event (Verhoeven et al., 2021). However, previous research has validated codes for stroke and MI in the UK Biobank, showing 80–90 % positive predictive value (Woodfield et al., 2015). Third, the use of codes both in primary and secondary position could lead to associations biased upwards. This might occur if the hospital visit has a non-CV ICD code in primary position and a CV code of interest as secondary. If the code in primary position is positively associated with air pollution, consequently the resulting association with the CV code will also be inflated. However, our sensitivity analysis using only codes occurring in primary position showed only partial changes in the association for the majority of the outcomes. We did not use primary codes in the main analysis only due to a low number of cases. Moreover, for the same reason, particularly for outcomes subtypes such as stroke and MI, the corresponding main associations displayed large uncertainty, leading to several non-statistically significant results. Finally, we only investigated one pollutant without accounting for other important pollutants, such as NO_2 and O_3, known to be associated with health outcomes.

5. Conclusions

Our study suggests that long-term exposure to $PM_{2.5}$ is associated with multiple cardiovascular outcomes. The strength of the associations

did not significantly vary using different exposure windows. Consistently with the current literature, we found increased associations at low levels of exposure for the majority of the outcomes highlighting the importance of estimating exposure–response functions in long-term air pollution analyses. Finally, we found that selecting specific diagnoses instead of broad outcomes definitions may be beneficial to identify more relevant health outcomes.

Funding

This work was supported by the Medical Research Council-UK (Grant ID: MR/Y003330/1), the European Union's Horizon 2020 Project Exhaustion (Grant ID: 820655), Nagasaki University "Doctoral Program for World-leading Innovative and Smart Education" for Global Health (WISE), KENKYU SHIDO KEIHI ("the Research Grant") and KYOIKU KENKYU SHIEN KEIHI ("the Stipend").

Declaration of Generative AI and AI-assisted technologies in the writing process

During the preparation of this work the author(s) used ChatGPT 3.5 to improve the language in some sentences in the Introduction and Discussion section. After using this tool/service, the author(s) reviewed and edited the content as needed and take(s) full responsibility for the content of the publication.

CRediT authorship contribution statement

Jacopo Vanoli: Writing – original draft, Visualization, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Jennifer K. Quint:** Writing – review & editing, Supervision, Methodology, Conceptualization. **Sanjay Rajagopalan:** Writing – review & editing, Methodology, Conceptualization. **Massimo Stafoggia:** Writing – review & editing, Methodology, Conceptualization. **Sadeer Al-Kindi:** Writing – review & editing, Methodology, Conceptualization. **Malcolm N. Mistry:** Writing – review & editing. **Pierre Masselot:** Writing – review & editing. **Arturo de la Cruz Libardi:** Writing – review & editing. **Chris Fook Sheng Ng:** Writing – review & editing. **Lina Madaniyazi:** Writing – review & editing, Supervision, Conceptualization. **Antonio Gasparrini:** Writing – review & editing, Supervision, Resources, Project administration, Methodology, Funding acquisition, Data curation, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The authors do not have permission to share data.

Acknowledgements

This research has been conducted using the UK Biobank Resource under application number 56431.

This work was part of a PhD project supported by the Nagasaki University "Doctoral Program for World-leading Innovative and Smart Education" for Global Health, KENKYU SHIDO KEIHI (the research grant) and KYOIKU KENKYU SHIEN KEIHI (the stipend).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.envint.2024.109011.

References

Alexeeff, S.E., Deosaransingh, K., Van Den Eeden, S., Schwartz, J., Liao, N.S., Sidney, S., 2023. Association of Long-term Exposure to Particulate Air Pollution With Cardiovascular Events in California. JAMA Netw. Open 6 (2), e230561. https://doi.org/10.1001/jamanetworkopen.2023.0561.

Alexeeff, S.E., Liao, N.S., Liu, X., VanDenEeden, S.K., Sidney, S., 2021. Long-Term PM2.5 Exposure and Risks of Ischemic Heart Disease and Stroke Events: Review and Meta-Analysis. Journal of the American Heart Association. 10 (1), e016890. https://doi.org/10.1161/JAHA.120.016890.

Al-Kindi, S.G., Brook, R.D., Biswal, S., Rajagopalan, S., 2020. Environmental determinants of cardiovascular disease: lessons learned from air pollution. Nat Rev Cardiol. 17 (10), 656–672. https://doi.org/10.1038/s41569-020-0371-2.

Andersen, P.K., Gill, R.D., 1982. Cox's Regression Model for Counting Processes: A Large Sample Study. The Annals of Statistics. 10 (4), 1100–1120.

Atkinson, R.W., Carey, I.M., Kent, A.J., van Staa, T.P., Anderson, H.R., Cook, D.G., 2013. Long-Term Exposure to Outdoor Air Pollution and Incidence of Cardiovascular Diseases. Epidemiology 24 (1), 44–53. https://doi.org/10.1097/EDE.0b013e318276ccb8.

Bai, L., Shin, S., Burnett, R.T., et al., 2019. Exposure to ambient air pollution and the incidence of congestive heart failure and acute myocardial infarction: A population-based study of 5.1 million Canadian adults living in Ontario. Environment International. 132, 105004. https://doi.org/10.1016/j.envint.2019.105004.

Barnard, J., Rubin, D.B., 1999. Small-Sample Degrees of Freedom with Multiple Imputation. Biometrika 86 (4), 948–955.

Bosco, E., Hsueh, L., McConeghy, K.W., Gravenstein, S., Saade, E., 2021. Major adverse cardiovascular event definitions used in observational analysis of administrative databases: a systematic review. BMC Med. Res. Method. 21 (1), 241. https://doi.org/10.1186/s12874-021-01440-5.

Cai, Y., Hodgson, S., Blangiardo, M., et al., 2018. Road traffic noise, air pollution and incident cardiovascular disease: A joint analysis of the HUNT, EPIC-Oxford and UK Biobank cohorts. Environ. Int. 114, 191–201. https://doi.org/10.1016/j.envint.2018.02.048.

Carlsen, H.K., Andersson, E.M., Molnár, P., et al., 2022. Incident cardiovascular disease and long-term exposure to source-specific air pollutants in a Swedish cohort. Environ. Res. 209, 112698. https://doi.org/10.1016/j.envres.2022.112698.

Cheema, K.M., Dicks, E., Pearson, J., Samani, N.J., 2022. Long-term trends in the epidemiology of cardiovascular diseases in the UK: insights from the British Heart Foundation statistical compendium. Cardiovasc. Res. 118 (10), 2267–2280. https://doi.org/10.1093/cvr/cvac053.

Chen, J., Braun, D., Christidis, T., et al., 2023. Long-Term Exposure to Low-Level PM2.5 and Mortality: Investigation of Heterogeneity by Harmonizing Analyses in Large Cohort Studies in Canada, United States, and Europe. Environ. Health Perspect. 131 (12), 127003. https://doi.org/10.1289/EHP12141.

Chen, C., Chen, H., Kaufman, J.S., Benmarhnia, T., 2024. Differential Participation, a Potential Cause of Spurious Associations in Observational Cohorts in Environmental Epidemiology. Epidemiology 35 (2), 174. https://doi.org/10.1097/EDE.0000000000001711.

Crouse, D.L., Peters, P.A., Hystad, P., et al., 2015. Ambient PM2.5, O3, and NO2 Exposures and Associations with Mortality over 16 Years of Follow-Up in the Canadian Census Health and Environment Cohort (CanCHEC). Environmental Health Perspectives. 123 (11), 1180–1186. https://doi.org/10.1289/ehp.1409276.

Crouse, D.L., Erickson, A.C., Christidis, T., et al., 2020. Evaluating the Sensitivity of PM2.5-Mortality Associations to the Spatial and Temporal Scale of Exposure Assessment. Epidemiology. 31 (2), 168–176. https://doi.org/10.1097/ede.0000000000001136.

Danesh Yazdi, M., Wang, Y., Di, Q., Zanobetti, A., Schwartz, J., 2019. Long-term exposure to PM2.5 and ozone and hospital admissions of Medicare participants in the Southeast USA. Environment International 130, 10487. https://doi.org/10.1016/j.envint.2019.05.073.

de Bont, J., Jaganathan, S., Dahlquist, M., Persson, Å., Stafoggia, M., Ljungman, P., 2022. Ambient air pollution and cardiovascular diseases: An umbrella review of systematic reviews and meta-analyses. J. Intern. Med. 291 (6), 779–800. https://doi.org/10.1111/joim.13467.

de la Cruz, L.A., Masselot, P., Schneider, R., et al., 2024. High resolution mapping of nitrogen dioxide and particulate matter in Great Britain (2003–2021) with multi-stage data reconstruction and ensemble machine learning methods. Atmos. Pollut. Res. 15 (11), 102284. https://doi.org/10.1016/j.apr.2024.102284.

Deo, S.V., Elgudin, Y., Motairek, I., et al., 2024. Air Pollution and Adverse Cardiovascular Events After Coronary Artery Bypass Grafting: A 10-Year Nationwide Study. JACC: Advances 3 (2), 100781. https://doi.org/10.1016/j.jacadv.2023.100781.

Di, Q., Wang, Y., Zanobetti, A., et al., 2017. Air Pollution and Mortality in the Medicare Population. N. Engl. J. Med. 376 (26), 2513–2522. https://doi.org/10.1056/NEJMoa1702747.

Fry, A., Littlejohns, T.J., Sudlow, C., et al., 2017. Comparison of Sociodemographic and Health-Related Characteristics of UK Biobank Participants With Those of the General Population. Am. J. Epidemiol. 186 (9), 1026–1034. https://doi.org/10.1093/aje/kwx246.

Gardner, B., Ling, F., Hopke, P.K., et al., 2014. Ambient fine particulate air pollution triggers ST-elevation myocardial infarction, but not non-ST elevation myocardial infarction: a case-crossover study. Part. Fibre Toxicol. 11 (1), 1. https://doi.org/10.1186/1743-8977-11-1.

General Lifestyle Survey: 2011. Office for national statistics; 2013.

Generalised Land Use Database Statistics for England 2005 (Enhanced Basemap). https://data.gov.uk/dataset/land_use_statistics_generalised_land_use_database.

How the NHS Cares for Patients with Heart Attack. Annual Public Report April 2014–March 2015. London: NICOR, 2017. NICOR.
Kelly FJ. *The Effects of Long-Term Exposure to Ambient Air Pollution on Cardiovascular Morbidity: Mechanistic Evidence*. COMEAP; 2019.
Kuźma, Ł., Dąbrowski, E.J., Kurasz, A., et al., 2024. Effect of air pollution exposure on risk of acute coronary syndromes in Poland: a nationwide population-based study (EP-PARTICLES study). The Lancet Regional Health - Europe. 41, 100910. https://doi.org/10.1016/j.lanepe.2024.100910.
Lefler, J., Higbee, J., Burnett, R., et al., 2019. Air pollution and mortality in a large, representative U.S. cohort: Multiple-pollutant analyses, and spatial and temporal decompositions. Environ. Health 18. https://doi.org/10.1186/s12940-019-0544-9.
Luo, H., Zhang, Q., Yu, K., Meng, X., Kan, H., Chen, R., 2022. Long-term exposure to ambient air pollution is a risk factor for trajectory of cardiometabolic multimorbidity: A prospective study in the UK Biobank. EBioMedicine 84. https://doi.org/10.1016/j.ebiom.2022.104282.
Mahdieh Danesh, Y., Yan, W., Qian, D., et al., 2021. Long-Term Association of Air Pollution and Hospital Admissions Among Medicare Participants Using a Doubly Robust Additive Model. Circulation 143 (16), 1584–1596. https://doi.org/10.1161/CIRCULATIONAHA.120.050252.
Management of Heart Attack: Analyses from the Myocardial Ischaemia National Audit Project (MINAP) and the National Audit of Percutaneous Coronary Intervention (NAPCI). NICOR; 2023.
Morton, D., Rowland, C., Wood, C., Meek, L., Marston, C., Smith, G., Simpson, I.C. 2011. Final Report for LCM2007 – the New UK Land Cover Map. CS Technical Report No 11/07 NERC/Centre for Ecology & Hydrology 108pp. (CEH Project Number: C03259). Based upon LCM2007 © NERC (CEH) 2011. Contains Ordnance Survey Data © Crown Copyright 2007. © Third Party Licensors.
Motairek, I., Deo, S.V., Elgudin, Y., et al., 2023. Particulate Matter Air Pollution and Long-Term Outcomes in Patients Undergoing Percutaneous Coronary Intervention. JACC: Advances 2 (3), 100285. https://doi.org/10.1016/j.jacadv.2023.100285.
Newby, D.E., Mannucci, P.M., Tell, G.S., et al., 2015. Expert position paper on air pollution and cardiovascular disease. Eur. Heart J. 36 (2), 83–93. https://doi.org/10.1093/eurheartj/ehu458.
Niu, Z., Liu, F., Yu, H., Wu, S., Xiang, H., 2021. Association between exposure to ambient air pollution and hospital admission, incidence, and mortality of stroke: an updated systematic review and meta-analysis of more than 23 million participants. Environ. Health Prev. Med. 26 (1), 15. https://doi.org/10.1186/s12199-021-00937-1.
Osborne, M.T., Abohashem, S., Naddaf, N., et al., 2023. The combined effect of air and transportation noise pollution on atherosclerotic inflammation and risk of cardiovascular disease events. J. Nucl. Cardiol. 30 (2), 665–679. https://doi.org/10.1007/s12350-022-03003-7.
Parra, K.L., Alexander, G.E., Raichlen, D.A., Klimentidis, Y.C., Furlong, M.A., 2022. Exposure to air pollution and risk of incident dementia in the UK Biobank. Environ. Res. 209, 112895. https://doi.org/10.1016/j.envres.2022.112895.
Pope, C.A., Muhlestein, J.B., Anderson, J.L., et al., 2015. Short-Term Exposure to Fine Particulate Matter Air Pollution Is Preferentially Associated With the Risk of ST-Segment Elevation Acute Coronary Events. J. Am. Heart Assoc. 4 (12), e002506. https://doi.org/10.1161/JAHA.115.002506.
Power, M.C., Lamichhane, A.P., Liao, D., et al., 2018. The Association of Long-Term Exposure to Particulate Matter Air Pollution with Brain MRI Findings: The ARIC Study. Environ. Health Perspect. 126 (2), 027009. https://doi.org/10.1289/EHP2152.
Pranata, R., Vania, R., Tondas, A.E., Setianto, B., Santoso, A., 2020. A time-to-event analysis on air pollutants with the risk of cardiovascular disease and mortality: A systematic review and meta-analysis of 84 cohort studies. J. Evid. Based Med. 13 (2), 102–115. https://doi.org/10.1111/jebm.12380.
Putter, H., van Houwelingen, H.C., 2017. Understanding Landmarking and Its Relation with Time-Dependent Cox Regression. Stat Biosci. 9 (2), 489–503. https://doi.org/10.1007/s12561-016-9157-9.
Sanjay, R., Sadeer, G.A.K., Robert, D.B., 2018. Air Pollution and Cardiovascular Disease: JACC State-of-the-Art Review. J. Am. Coll. Cardiol. 72 (17), 2054–2070. https://doi.org/10.1016/j.jacc.2018.07.099.
Shin, S., Burnett, R.T., Kwong, J.C., et al., 2019. Ambient Air Pollution and the Risk of Atrial Fibrillation and Stroke: A Population-Based Cohort Study. Environ. Health Perspect. 127 (8), 087009. https://doi.org/10.1289/EHP4883.
Stafoggia, M., Oftedal, B., Chen, J., et al., 2022. Long-term exposure to low ambient air pollution concentrations and mortality among 28 million people: results from seven large European cohorts within the ELAPSE project. Lancet Planet Health. 6 (1), e9–e18. https://doi.org/10.1016/s2542-5196(21)00277-1.
STEMI and NSTEMI: are they so different? 1 year outcomes in acute myocardial infarction as defined by the ESC/ACC definition (the OPERA registry)† | European Heart Journal | Oxford Academic. Accessed August 6, 2024. https://academic.oup.com/eurheartj/article/28/12/1409/2887253?login=false.
Sudlow, C., Gallacher, J., Allen, N., et al., 2015. UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. PLoSMedicine. 12 (3), e1001779. https://doi.org/10.1371/journal.pmed.1001779.
Vaduganathan, M., Mensah, G.A., Turco, J.V., Fuster, V., Roth, G.A., 2022. The Global Burden of Cardiovascular Diseases and Risk. J. Am. College Cardiol. 80 (25), 2361–2371. https://doi.org/10.1016/j.jacc.2022.11.005.
Vanoli, J., Mistry, M.N., De La Cruz Libardi, A., Masselot, P., Schneider, R., Ng, C.F.S., Madaniyazi, L., Gasparrini, A., 2024. Reconstructing individual-level exposures in cohort analyses of environmental risks: an example with the UK Biobank. J Expo Sci Environ Epidemiol 1–6. https://doi.org/10.1038/s41370-023-00635-w.
Verhoeven, J.I., Allach, Y., Vaartjes, I.C.H., Klijn, C.J.M., de Leeuw, F.E., 2021. Ambient air pollution and the risk of ischaemic and haemorrhagic stroke. The Lancet Planetary Health. 5 (8), e542–e552. https://doi.org/10.1016/S2542-5196(21)00145-5.
Wilker, E.H., Mostofsky, E., Fossa, A., et al., 2018. Ambient Pollutants and Spontaneous Intracerebral Hemorrhage in Greater Boston. Stroke 49 (11), 2764–2766. https://doi.org/10.1161/STROKEAHA.118.023128.
Wolf, K., Hoffmann, B., Andersen, Z.J., et al., 2021. Long-term exposure to low-level ambient air pollution and incidence of stroke and coronary heart disease: a pooled analysis of six European cohorts within the ELAPSE project. The Lancet Planetary Health. 5 (9), e620–e632. https://doi.org/10.1016/S2542-5196(21)00195-9.
Woodfield, R., Grant, I., Sudlow, C.L.M., 2015. Accuracy of Electronic Health Record Data for Identifying Stroke Cases in Large-Scale Epidemiological Studies: A Systematic Review from the UK Biobank Stroke Outcomes Group. PLoS One 10 (10), e0140533. https://doi.org/10.1371/journal.pone.0140533.
Zhang, S., Chen, L., Qian, Z. (Min), et al., 2024. Associations between air pollution and the risk of first admission and multiple readmissions for cardiovascular diseases. Heart 110 (5), 337–345. https://doi.org/10.1136/heartjnl-2023-322682.
Zhu, W., Cai, J., Hu, Y., et al., 2021. Long-term exposure to fine particulate matter relates with incident myocardial infarction (MI) risks and post-MI mortality: A meta-analysis. Chemosphere 267, 128903. https://doi.org/10.1016/j.chemosphere.2020.128903.

~

Association between long-term exposure to low ambient $PM_{2.5}$ and cardiovascular hospital admissions: a UK Biobank study

Appendix

Authors: Jacopo Vanoli[1,2], Jennifer K Quint[3], Sanjay Rajagopalan[4], Massimo Stafoggia[5], Sadeer Al-Kindi[6], Malcolm Mistry[2,7], Pierre Masselot[2], Arturo de la Cruz Libardi[2], Chris Fook Sheng Ng[8], Lina Madaniyazi[1], Antonio Gasparrini[2]

[1]School of Tropical Medicine and Global Health, Nagasaki University, Nagasaki, Japan, [2] Environment & Health Modelling (EHM) Lab, Department of Public Health Environments and Society, London School of Hygiene & Tropical Medicine, London, UK, [3]School of Public Health, Imperial College London, London, [4]Harrington Heart and Vascular Institute, University Hospitals Cleveland Medical Center, Cleveland, OH, 44106, USA, [5]Department of Epidemiology, Lazio Region Health Service ASL ROMA 1, Rome, Italy, [6]Center for Health and Nature, Houston Methodist, Houston, Texas, United States, [7]Department of Economics, Ca' Foscari University of Venice, Venice, Italy, [8]Department of Global Health Policy, Graduate School of Medicine, The University of Tokyo, Tokyo, Japan

Table S1: Hazard ratios (HRs, with 95% confidence intervals) of cardiovascular outcomes with ICD in primary position associated with an increase of 5 μg/m^3 in $PM_{2.5}$ in the UKB cohort, for combinations of length of exposure windows and confounding control.

Outcome	exposure window	Model 1	Model 2
5-point MACE	lag0	1.13 (1.03-1.24)	1.09 (0.98-1.22)
	lag02	1.14 (1.03-1.26)	1.10 (0.98-1.24)
	lag04	1.13 (1.01-1.25)	1.08 (0.95-1.22)
Myocardial Infarction (MI)			
Acute	lag0	0.93 (0.80-1.09)	0.99 (0.83-1.18)
	lag02	0.91 (0.77-1.07)	0.96 (0.79-1.16)
	lag04	0.91 (0.77-1.08)	0.97 (0.79-1.18)
STEMI	lag0	0.83 (0.65-1.05)	0.83 (0.64-1.09)
	lag02	0.86 (0.67-1.11)	0.88 (0.66-1.18)
	lag04	0.90 (0.69-1.17)	0.94 (0.69-1.27)
NSTEMI	lag0	1.12 (0.87-1.46)	1.39 (1.02-1.89)
	lag02	1.00 (0.76-1.32)	1.22 (0.87-1.70)
	lag04	0.98 (0.74-1.29)	1.17 (0.84-1.64)
Cerebrovascular disease and stroke			
Intracerebral stroke	lag0	1.64 (1.05-2.54)	1.63 (0.98-2.72)
	lag02	1.81 (1.14-2.88)	1.89 (1.09-3.29)
	lag04	1.79 (1.11-2.89)	1.87 (1.05-3.31)
Ischaemic stroke	lag0	1.22 (1.02-1.47)	1.14 (0.92-1.40)
	lag02	1.25 (1.03-1.52)	1.15 (0.92-1.45)

	lag04	1.20 (0.98-1.47)	1.08 (0.86-1.37)
Subarachnoid stroke	lag0	1.14 (0.71-1.84)	1.06 (0.62-1.81)
	lag02	1.15 (0.69-1.92)	1.05 (0.58-1.91)
	lag04	1.12 (0.65-1.92)	1.00 (0.53-1.88)
Other outcomes			
Heart failure	lag0	1.52 (1.14-2.03)	1.26 (0.91-1.74)
	lag02	1.59 (1.17-2.16)	1.28 (0.90-1.83)
	lag04	1.66 (1.21-2.28)	1.34 (0.93-1.93)
Atrial fibrillation and flutter	lag0	2.85 (1.32-6.16)	3.49 (1.46-8.38)
	lag02	3.70 (1.65-8.28)	5.17 (2.05-13.08)
	lag04	3.88 (1.67-9.01)	5.56 (2.10-14.72)
Cardiac arrest	lag0	1.12 (0.98-1.28)	1.15 (0.98-1.34)
	lag02	1.11 (0.96-1.28)	1.14 (0.96-1.35)
	lag04	1.12 (0.96-1.30)	1.15 (0.96-1.37)

Table S2: Hazard ratios (HRs, with 95% confidence intervals) of cardiovascular outcomes associated with an increase of 5 μg/m^3 in $PM_{2.5}$ in the UKB cohort with follow-up until 2019, for combinations of length of exposure windows and confounding control.

Outcome	**exposure window**	**Model 1**	**Model 2**
5-point MACE	lag0	1.15 (1.06-1.26)	1.11 (1.00-1.23)
	lag02	1.17 (1.06-1.28)	1.12 (1.00-1.24)
	lag04	1.16 (1.05-1.28)	1.10 (0.98-1.23)
Myocardial Infarction (MI)			
Acute	lag0	0.96 (0.83-1.11)	1.06 (0.89-1.26)
	lag02	0.94 (0.81-1.11)	1.04 (0.86-1.25)
	lag04	0.95 (0.80-1.12)	1.05 (0.86-1.28)
STEMI	lag0	0.93 (0.74-1.17)	1.04 (0.80-1.36)
	lag02	0.93 (0.73-1.19)	1.06 (0.80-1.42)
	lag04	0.95 (0.74-1.23)	1.11 (0.82-1.51)
NSTEMI	lag0	1.06 (0.83-1.36)	1.39 (1.02-1.88)
	lag02	1.00 (0.77-1.29)	1.27 (0.92-1.75)
	lag04	1.00 (0.77-1.31)	1.28 (0.92-1.77)
Cerebrovascular disease and stroke			
Intracerebral stroke	lag0	1.72 (1.16-2.57)	1.65 (1.03-2.64)
	lag02	1.89 (1.25-2.88)	1.88 (1.14-3.09)
	lag04	1.90 (1.23-2.94)	1.87 (1.11-3.15)
Ischaemic stroke	lag0	1.07 (0.90-1.29)	0.99 (0.80-1.22)
	lag02	1.11 (0.92-1.34)	1.02 (0.82-1.27)

	lag04	1.07 (0.88-1.31)	0.97 (0.77-1.22)
Subarachnoid stroke	lag0	1.01 (0.64-1.58)	0.91 (0.55-1.53)
	lag02	1.07 (0.66-1.73)	0.98 (0.56-1.71)
	lag04	1.12 (0.67-1.87)	1.03 (0.57-1.88)
Other outcomes			
Heart failure	lag0	1.37 (1.17-1.59)	1.22 (1.01-1.46)
	lag02	1.37 (1.17-1.61)	1.20 (0.99-1.45)
	lag04	1.41 (1.19-1.66)	1.22 (1.00-1.49)
Atrial fibrillation and flutter	lag0	1.11 (1.00-1.23)	1.10 (0.98-1.24)
	lag02	1.13 (1.02-1.27)	1.13 (0.99-1.28)
	lag04	1.12 (1.00-1.25)	1.10 (0.96-1.26)
Cardiac arrest	lag0	1.11 (1.00-1.23)	1.09 (0.97-1.22)
	lag02	1.13 (1.01-1.25)	1.11 (0.98-1.25)
	lag04	1.11 (0.99-1.24)	1.08 (0.95-1.23)

Table S3: Complete version of table 3 including all the length of the exposure window. Hazard ratios (HRs, with 95% confidence intervals) of cardiovascular outcomes in exposure subset analysis in the UKB cohort. Results are for an increase of 5 μg/m^3 in $PM_{2.5}$ using fully adjusted models (Model 2).

Outcome	**exposure window**	**Exposure subset**		
		<= 10	**<= 12**	**Full analysis**
5-point MACE	lag0	1.23 (1.03-1.46)	1.16 (1.03-1.31)	1.13 (1.02-1.24)
	lag02	1.21 (0.98-1.48)	1.20 (1.05-1.37)	1.15 (1.03-1.28)
	lag04	1.05 (0.84-1.32)	1.16 (1.01-1.34)	1.12 (1.00-1.26)
Myocardial Infarction (MI)				
Acute	lag0	1.45 (1.11-1.90)	1.04 (0.76-1.41)	1.07 (0.90-1.27)
	lag02	1.34 (0.97-1.84)	1.01 (0.71-1.45)	1.06 (0.87-1.28)
	lag04	1.27 (0.90-1.81)	0.78 (0.53-1.17)	1.06 (0.87-1.29)
STEMI	lag0	1.26 (1.01-1.56)	1.08 (0.84-1.38)	0.98 (0.75-1.28)
	lag02	1.19 (0.94-1.51)	1.01 (0.77-1.34)	1.03 (0.77-1.38)
	lag04	1.20 (0.94-1.53)	1.07 (0.81-1.43)	1.09 (0.80-1.49)
NSTEMI	lag0	1.19 (0.79-1.79)	0.88 (0.39-1.97)	1.52 (1.12-2.07)
	lag02	1.26 (0.77-2.08)	0.89 (0.33-2.36)	1.35 (0.97-1.88)
	lag04	1.19 (0.68-2.07)	0.75 (0.25-2.22)	1.32 (0.95-1.84)
Cerebrovascular disease and stroke				
Intracerebral stroke	lag0	1.94 (1.31-2.87)	1.14 (0.97-1.33)	1.74 (1.09-2.78)
	lag02	1.68 (1.06-2.66)	1.07 (0.88-1.30)	1.91 (1.15-3.17)
	lag04	1.70 (1.01-2.86)	1.10 (0.89-1.36)	1.94 (1.15-3.29)

Ischaemic stroke	lag0	1.63 (1.17-2.26)	1.08 (0.95-1.22)	1.07 (0.87-1.32)
	lag02	1.48 (1.03-2.12)	1.01 (0.88-1.16)	1.08 (0.86-1.36)
	lag04	1.45 (1.00-2.10)	0.98 (0.85-1.14)	1.01 (0.80-1.28)
Subarachnoid stroke	lag0	1.04 (0.85-1.29)	1.28 (1.00-1.65)	0.94 (0.56-1.56)
	lag02	1.11 (0.86-1.42)	1.32 (0.97-1.78)	0.98 (0.56-1.74)
	lag04	0.98 (0.74-1.29)	1.33 (0.95-1.87)	1.03 (0.56-1.89)
Other outcomes				
Heart failure	lag0	1.62 (0.80-3.30)	1.06 (0.93-1.21)	1.19 (1.00-1.42)
	lag02	2.06 (0.86-4.95)	1.07 (0.92-1.25)	1.21 (1.00-1.48)
	lag04	1.51 (0.58-3.93)	1.13 (0.95-1.35)	1.22 (1.00-1.50)
Atrial fibrillation and flutter	lag0	0.95 (0.53-1.69)	1.08 (0.92-1.27)	1.39 (0.94-2.05)
	lag02	0.85 (0.43-1.69)	1.09 (0.90-1.33)	1.38 (0.90-2.12)
	lag04	0.58 (0.27-1.22)	1.22 (0.98-1.52)	1.26 (0.81-1.95)
Cardiac arrest	lag0	1.38 (0.86-2.21)	1.06 (0.93-1.21)	1.09 (0.98-1.22)
	lag02	1.17 (0.69-1.98)	1.12 (0.97-1.30)	1.15 (1.02-1.29)
	lag04	0.95 (0.55-1.64)	1.08 (0.93-1.25)	1.16 (1.03-1.31)

Table S4: Hazard ratios (HRs, with 95% confidence intervals) of three MACE outcomes definitions associated with an increase of 5 μg/m^3 in $PM_{2.5}$ in the UKB cohort, for combinations of length of exposure windows and confounding control. *Results for 5-point MACE are the same as Table 2.

Outcome	**exposure window**	**Model 1**	**Model 2**
3-point MACE	lag0	1.11 (1.00-1.23)	1.10 (0.98-1.24)
	lag02	1.13 (1.02-1.27)	1.13 (0.99-1.28)
	lag04	1.12 (1.00-1.25)	1.10 (0.96-1.26)
4-point MACE	lag0	1.11 (1.00-1.23)	1.09 (0.97-1.22)
	lag02	1.13 (1.01-1.25)	1.11 (0.98-1.25)
	lag04	1.11 (0.99-1.24)	1.08 (0.95-1.23)
5-point MACE*	lag0	1.17 (1.08-1.28)	1.13 (1.02-1.24)
	lag02	1.20 (1.09-1.32)	1.15 (1.03-1.28)
	lag04	1.18 (1.07-1.31)	1.12 (1.00-1.26)

Table S5: Hazard ratios (HRs, with 95% confidence intervals) of cardiovascular outcomes associated with an increase of the IQR (3.7 μg/m^3) in PM2.5 in the UKB cohort for combinations of length of exposure windows and confounding control.

Outcome	exposure window	Model 1	Model 2
5-point MACE	lag0	1.13 (1.06-1.20)	1.09 (1.01-1.18)
	lag02	1.14 (1.07-1.23)	1.11 (1.02-1.20)
	lag04	1.13 (1.05-1.22)	1.09 (1.00-1.18)
Myocardial Infarction (MI)			
Acute	lag0	0.98 (0.88-1.10)	1.05 (0.93-1.20)
	lag02	0.97 (0.86-1.09)	1.04 (0.91-1.20)
	lag04	0.97 (0.85-1.09)	1.04 (0.90-1.21)
STEMI	lag0	0.92 (0.77-1.09)	1.00 (0.82-1.21)
	lag02	0.93 (0.78-1.12)	1.04 (0.84-1.29)
	lag04	0.95 (0.79-1.16)	1.08 (0.86-1.36)
NSTEMI	lag0	1.12 (0.93-1.36)	1.37 (1.09-1.72)
	lag02	1.03 (0.85-1.26)	1.26 (0.99-1.60)
	lag04	1.02 (0.84-1.25)	1.24 (0.97-1.58)
Cerebrovascular disease and stroke			
Intracerebral stroke	lag0	1.54 (1.14-2.08)	1.49 (1.05-2.11)
	lag02	1.62 (1.18-2.22)	1.60 (1.09-2.33)
	lag04	1.63 (1.18-2.27)	1.62 (1.10-2.39)
Ischaemic stroke	lag0	1.10 (0.96-1.26)	1.06 (0.90-1.23)
	lag02	1.12 (0.97-1.29)	1.07 (0.90-1.26)
	lag04	1.09 (0.94-1.26)	1.02 (0.86-1.21)
Subarachnoid stroke	lag0	1.01 (0.72-1.42)	0.94 (0.65-1.38)
	lag02	1.05 (0.73-1.52)	0.98 (0.64-1.48)
	lag04	1.09 (0.74-1.59)	1.01 (0.64-1.58)
Other outcomes			
Heart failure	lag0	1.25 (1.11-1.40)	1.14 (1.00-1.30)
	lag02	1.28 (1.13-1.45)	1.16 (1.00-1.34)
	lag04	1.29 (1.14-1.47)	1.16 (1.00-1.35)
Atrial fibrillation and flutter	lag0	1.30 (1.01-1.68)	1.33 (0.99-1.78)
	lag02	1.31 (1.00-1.71)	1.34 (0.97-1.84)
	lag04	1.24 (0.94-1.64)	1.25 (0.90-1.73)
Cardiac arrest	lag0	1.10 (1.02-1.17)	1.07 (0.99-1.16)
	lag02	1.13 (1.05-1.22)	1.11 (1.02-1.21)
	lag04	1.14 (1.05-1.23)	1.11 (1.02-1.22)

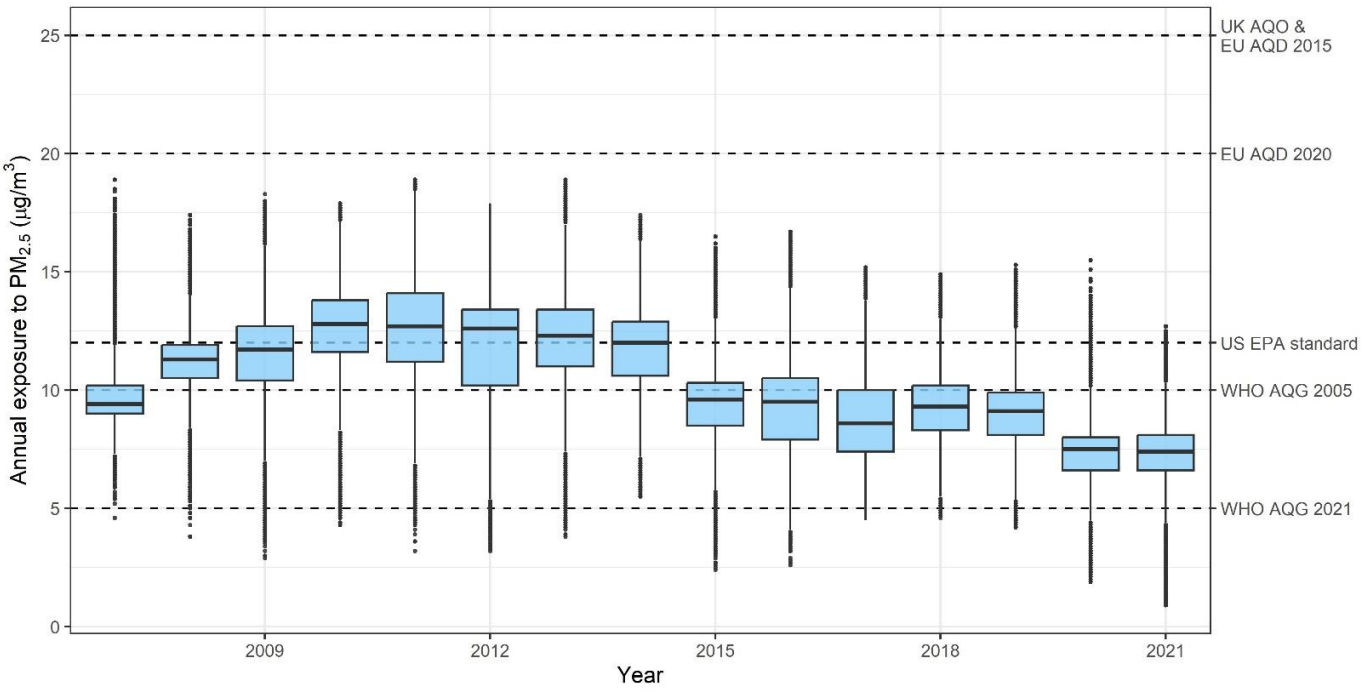
Table S6: Hazard ratios (HRs, with 95% confidence intervals) of cardiovascular outcomes associated with an increase of 5 μg/m^3 in PM2.5 in the UKB cohort excluding the *wash-out* period from enrolment until 2013 for combinations of length of exposure windows and confounding control.

Outcome	**exposure window**	**Model 1**	**Model 2**
5-point MACE	lag0	1.24 (1.12-1.38)	1.20 (1.06-1.36)
	lag02	1.26 (1.13-1.41)	1.22 (1.07-1.38)
	lag04	1.25 (1.12-1.39)	1.20 (1.05-1.36)
Myocardial Infarction (MI)			
Acute	lag0	1.15 (0.95-1.39)	1.31 (1.05-1.63)
	lag02	1.12 (0.92-1.36)	1.29 (1.02-1.63)
	lag04	1.12 (0.93-1.37)	1.30 (1.03-1.64)
STEMI	lag0	1.13 (0.85-1.51)	1.28 (0.91-1.79)
	lag02	1.19 (0.88-1.61)	1.40 (0.97-2.00)
	lag04	1.22 (0.90-1.66)	1.45 (1.01-2.09)
NSTEMI	lag0	1.14 (0.87-1.49)	1.46 (1.06-2.02)
	lag02	1.03 (0.78-1.36)	1.31 (0.93-1.84)
	lag04	1.03 (0.78-1.35)	1.29 (0.92-1.80)
Cerebrovascular disease and stroke			
Intracerebral stroke	lag0	1.68 (1.05-2.68)	1.77 (1.03-3.06)
	lag02	1.76 (1.08-2.86)	1.92 (1.07-3.43)
	lag04	1.75 (1.07-2.84)	1.91 (1.06-3.42)
Ischaemic stroke	lag0	1.12 (0.90-1.40)	1.07 (0.83-1.37)
	lag02	1.11 (0.89-1.39)	1.04 (0.80-1.36)
	lag04	1.09 (0.87-1.36)	1.01 (0.77-1.31)
Subarachnoid stroke	lag0	0.89 (0.50-1.57)	0.84 (0.44-1.60)
	lag02	0.97 (0.53-1.77)	0.94 (0.47-1.89)
	lag04	1.02 (0.56-1.88)	1.00 (0.49-2.05)
Other outcomes			
Heart failure	lag0	1.41 (1.18-1.69)	1.23 (1.00-1.51)
	lag02	1.51 (1.25-1.81)	1.31 (1.05-1.64)
	lag04	1.50 (1.25-1.81)	1.30 (1.05-1.63)
Atrial fibrillation and flutter	lag0	1.51 (1.01-2.24)	1.55 (0.98-2.46)
	lag02	1.50 (0.99-2.27)	1.54 (0.94-2.51)
	lag04	1.40 (0.92-2.12)	1.40 (0.86-2.28)
Cardiac arrest	lag0	1.17 (1.05-1.31)	1.15 (1.01-1.30)
	lag02	1.22 (1.08-1.36)	1.20 (1.05-1.38)
	lag04	1.23 (1.09-1.38)	1.21 (1.06-1.39)

Table S7: Correlations between different exposure windows.

exposure windows	lag0	lag02	lag04
lag0	1		
lag02	0.93	1	
lag04	0.84	0.96	1

Figure S1: Distribution of annual average exposure to $PM_{2.5}$ across participants of the UKB cohort in the period 2003-2021, with limits corresponding to air quality guidelines/directives (AQG and AQD) from the World Health Organization (WHO), European Union (EU), and the United Kingdom (UK).



Chapter 12

Research article 4

Confounding mechanisms and adjustment strategies in air pollution epidemiology: an empirical assessment with the UK Biobank cohort

London School of Hygiene & Tropical Medicine
Keppel Street, London WC1E 7HT
T: +44 (0)20 7299 4646
F: +44 (0)20 7299 4656
www.lshtm.ac.uk

国立大学法人
長崎大学
NAGASAKI UNIVERSITY

RESEARCH PAPER COVER SHEET

Please note that a cover sheet must be completed for each research paper included within a thesis.

SECTION A – Student Details

Nagasaki Student No	59720005	**Title**	Confounding in air pollution epidemiology: an empirical assessment with the UK Biobank cohort
LSHTM Student ID No	2004062		
First Name(s)	Jacopo		
Surname/Family Name	Vanoli		
Thesis Title	Assessment of the effect of air pollution on the UK Biobank cohort		
Nagasaki Supervisor(s)	Lina Madaniyazi		
LSHTM Supervisor(s)	Antonio Gasparrini		

If the Research Paper has previously been published please complete Section B, if not please move to Section C.

SECTION B – Paper already published

Where was the work published?			
When was the work published?			
If the work was published prior to registration for your research degree, give a brief rationale for its inclusion			
Have you retained the copyright for the work?*	Choose an item.	Was the work subject to academic peer review?	Choose an item.

*If yes, please attach evidence of retention. If no, or if the work is being included in its published format, please attach evidence of permission from the copyright holder (publisher or other author) to include this work.

SECTION C – Prepared for publication, but not yet published

Where is the work intended to be published?	International Journal of Epidemiology

Please list the paper's authors in the intended authorship order:	Jacopo Vanoli, Lina Madaniyazi, Massimo Stafoggia, Chris Fook Sheng Ng, Antonio Gasparrini
Stage of publication	**Not yet submitted**

SECTION D – Multi-authored work

For multi-authored work, give full details of your role in the research included in the paper and in the preparation of the paper. (Attach a further sheet if necessary)	**I took the lead in carrying out the analysis, structuring and writing up the paper**

SECTION E – Names and affiliations of co-author(s)

Please list all the co-authors' names and their affiliations.

Jacopo Vanoli[1,2], Lina Madaniyazi[1], Massimo Stafoggia[3], Chris Fook Sheng Ng[4], Antonio Gasparrini[2]
[1]School of Tropical Medicine and Global Health, Nagasaki University, Nagasaki, Japan, [2]Environment & Health Modelling (EHM) Lab, Department of Public Health Environments and Society, London School of Hygiene & Tropical Medicine, London, UK, [3]Department of Epidemiology, Lazio Region Health Service ASL ROMA 1, Rome, Italy, [4]Department of Global Health Policy, Graduate School of Medicine, The University of Tokyo, Tokyo, Japan

SECTION F

I confirm that all co-authors have agreed that the above paper will be included in my PhD thesis.

Student Signature	[redacted]
Date	9/5/2024

LSHTM Supervisor Signature	[redacted]
Date	05/09/2024

Nagasaki University Supervisor Signature	[redacted]
Date	05/09/2024

Please submit this form to: tmghadmin@ml.nagasaki-u.ac.jp

Confounding mechanisms and adjustment strategies in air pollution epidemiology: a case study assessment with the UK Biobank cohort

Authors: Jacopo Vanoli[1,2], Lina Madaniyazi[1], Massimo Stafoggia[3], Chris Fook Sheng Ng[4], Antonio Gasparrini[2]

[1]School of Tropical Medicine and Global Health, Nagasaki University, Nagasaki, Japan, [2]Environment & Health Modelling (EHM) Lab, Department of Public Health Environments and Society, London School of Hygiene & Tropical Medicine, London, UK, [3]Department of Epidemiology, Lazio Region Health Service ASL ROMA 1, Rome, Italy, [4]Department of Global Health Policy, Graduate School of Medicine, The University of Tokyo, Tokyo, Japan

Corresponding author: Jacopo Vanoli, jacopo.vanoli@lshtm.ac.uk

Abstract

Background

Cohort studies are instrumental in examining long-term risks associated with environmental exposures. However, the reliability of their estimates requires appropriate control for various confounding mechanisms. In this contribution, we elucidate and discuss different confounding processes by investigating the relationship between fine particulate matter ($PM_{2.5}$) exposure and mortality in a UK-based cohort.

Methods

We analysed data from half a million adults in the UK Biobank linked with annual individual-level exposure data and followed up during the period 2006-2021. The assessment focused on confounding related to spatial and temporal patterns as well as due to measurable variables, including both contextual and individual-level factors. We performed a comprehensive evaluation consisting of descriptive analyses, specification and interpretation of direct acyclic graphs (DAGs), and comparison of results from Cox proportional hazard models with different adjustment strategies.

Results

We found correlations between both $PM_{2.5}$ exposure and mortality rates across time, geographical areas, and categories of measurable variables. The analysis of the DAG indicated complex causal pathways and the need to adjust for a wide set of potential confounders. Results from regression models confirm these patterns: the fully-adjusted model estimated a hazard ratio (HR) of 1.25 (95%CI: 1.06-1.49) per 10 μg/m³ increments in $PM_{2.5}$, but the association reversed to 0.82 (0.76-0.87) when excluding control for recruitment centre, suggesting strong spatial confounding. Calendar time showed stronger confounding effects compared to age. Area-level socio-economic indicators such as deprivation were more important than individual-level counterparts, while lack of control for lifestyle factors such as smoking, alcohol consumption, and waist-to-hip ratio led to a noticeable overestimation.

Conclusions

This case-study illustration identified various confounding mechanisms in cohort studies on long-term risks of environmental exposures. These patterns require specific adjustment strategies to control for spatial differences, temporal trends, and direct confounding from a range of contextual and individual-level variables.

Key messages

- The assessment of long-term risks associated with environmental exposures such as air pollution relies on observational analyses of cohort datasets that are prone to various confounding mechanisms.
- In this contribution, we examine and characterise confounding processes related to both spatial and temporal patterns as well as linked to both contextual and individual-level measurable factors. We offer a comprehensive overview based on theoretical considerations and empirical findings, using as a case study an analysis of the UK Biobank cohort
- The analysis of direct acyclic graphs (DAGs) and substantive results suggest various potential confounding mechanisms, particularly those related to unmeasured spatial differences in exposure levels and baseline risks, as well as individual-level lifestyle factors.
- These findings emphasise the potential impact of confounding in cohort studies on long-term risks of air pollution and the importance of devising appropriate epidemiological methods to control for it.

Introduction

Epidemiological evidence on long-term risks of environmental stressors, such as air pollution, comes mostly from population-based cohort studies, with landmark publications that have contributed to establishing causal links between exposure to several pollutants and various health effects, primarily mortality[1,2]. Such observational studies rely on the follow-up of large samples of individuals over extended periods, and they apply time-to-event designs based on between-subject comparisons to estimate potentially complex temporal relationships. Therefore, they require specific design choices and covariate adjustment to avoid confounding due to differential exposure and health risks across individuals, as well as over space and time[3–5].

In this article, we discuss different confounding processes that can affect cohort studies on environmental risk factors, differentiating them based on their source. Specifically, we classify them into spatial and temporal mechanisms, as well as those related to measurable risk factors, the latter distinguished between contextual and individual-level variables. Spatial confounding concerns variations in exposure and health risks determined by (mostly unmeasured) area-level characteristics. In this case, the comparison of subjects living in different areas is prone to confounding due to differential baseline risks related to environmental exposures[4]. Similarly, temporal confounding can manifest itself through collinear patterns of exposure and outcome across time. This is especially relevant in studies on long-term risks, given the steadily decreasing trends in air pollution levels and the changing patterns in disease occurrence over time within a given population[6]. Lastly, confounding can be related to measurable variables, including a wide spectrum of risk factors such as socio-economic indices, environmental conditions, as well as lifestyle and other personal characteristics.

Control for residual spatial and temporal confounding is usually achieved by implementing specific design and modelling strategies, such as stratification, selection of the time axis, and inclusion of variables such as region, age, and calendar time[4,7]. Differently, confounding patterns related to putative risk factors are more difficult to conceptualise and untangle. Their control can be achieved by the inclusion of measurable variables in regression models, and decisions about their selection usually require additional assumptions on the potential causal pathways, facilitated by the application of specific methods such as directed acyclic graphs (DAGs)[8]. Previous works have discussed methodological and practical issues related to confounding in cohort studies on air pollution[4,9,10]. However, several questions remain unaddressed, such as the appropriate strategies to account for spatial and temporal differences in time-to-event analysis, or the role of individual-level factors, in particular those related to lifestyle aspects that are rarely available in large administrative health databases.

In this contribution, we examine confounding patterns in the relationship between long-term exposure to air pollution and mortality within the UK Biobank cohort. This large database offers rich individual-level information and detailed exposure histories, making it an ideal case study to inspect potential confounding effects. Through theoretical considerations, descriptive analyses, and empirical estimates of health risk associations, we explore various confounding mechanisms and model adjustment strategies.

Data and Methods

UK Biobank

The UK Biobank (UKB) is a large biomedical database that includes approximately half a million participants[11]. Recruitment occurred between 2006 and 2010 for adults aged 40-69 residing within 10 miles of one of 22 assessment centres spread across the UK. The catchment areas of the assessment centres were chosen based on sufficient proximity to highly-populated regions. The participants underwent a first in-person visit during which they completed several questionnaires regarding their personal characteristics, lifestyles and medical history. Anthropometric measurements and blood samples were also taken. Specific details regarding the UKB database can be found on the showcase website (https://biobank.ndph.ox.ac.uk/showcase/).

Air pollution exposure

Individual-level exposure to outdoor fine particulate matter ($PM_{2.5}$) was assigned to each participant accounting for the residential history across the follow-up. The original $PM_{2.5}$ data were defined at daily level and predicted on a 1-km grid across the UK in the period 2003-2021 using a hybrid spatio-temporal machine learning model[12]. The residential data were available in the UKB database, including periods and geocoded locations with 100m rounding. The linkage process involved constructing daily exposure series for each residential location and then composing subject-specific exposure profiles by linking the daily series for corresponding residential periods, finally aggregating the data in annual averages. The process has been described in a previous publication[13].

Outcomes

At the time of enrolment, the participants consented to access their electronic health records. This data includes health information routinely collected from the NHS system as well as national health registries. Cause and date of death were extracted by the death national registry. The outcome events were defined as deaths due to non-accidental causes (ICD10: A00-R99).

Selection of confounders

We selected the set of confounders based on substantive knowledge of their role in the relationship between outdoor $PM_{2.5}$ concentrations and mortality, and theoretical results from the application of DAGs. Contextual features were represented by the assessment centre and variables defined at the residence at recruitment time, including area-level deprivation measured by the Townsend Deprivation Index (TDI), the percentage of greenspace within a 1000m buffer, and urban-rural classification. Individual-level confounders were separated into socio-economic factors (ethnic background, education level, household income, and employment status) and physical and lifestyle characteristics (smoking status and intensity, alcohol intake, waist-to-hip ratio, physical activity, and living alone). All of them were defined at baseline. Details on the covariate definitions are provided in the Appendix.

The assumptions on the causal paths assumed between outdoor $PM_{2.5}$ exposure, the other risk factors listed above, and mortality are represented as a DAG of Figure 1. The diagram includes assessment centre and the three measurable area-level variables, plus individual-level risk factors grouped in socio-economic and lifestyle characteristics, respectively. In addition, the graph comprises unobserved variables, specifically a group representing unmeasured centre-level characteristics and personal $PM_{2.5}$ exposure.

Statistical analysis

The analysis was based on a design strategy and model selection described in detail in previous work[14]. We applied a Cox proportional hazard model with time-varying exposures. The exposure index was defined at individual level as the time-varying average of $PM_{2.5}$ in the eight calendar years before the event. We assumed a linear exposure-response relationship between exposure and outcome. The end of follow-up was determined either by subject's death, loss to follow-up, or the administrative end of mortality linkage (31 December 2022).

In our main (full) model, we used calendar time as the time axis, we stratified by indicators of sex, assessment centre, and month of birth, and finally, we adjusted directly for all the other covariates listed in the previous section. We then defined alternative adjustment strategies to examine potential confounding mechanisms, specifically by excluding confounders or stratifying variables from the model (individually or in groups), and by varying time axes and control of trends.

We conducted a sensitivity analysis for the inclusion of the assessment centre in the statistical model using different specifications, following a previous work[10]. Specifically, we used dummy variables or random effects with a gamma distribution or normal distribution.

Estimates of the associations were reported as hazard ratios (HRs) for non-accidental mortality per 10 μg/m³ increments in $PM_{2.5}$, with 95% confidence intervals (CI). Missing values in the baseline covariates were imputed using multiple imputations by chained equation (MICE). For simplicity, we used the results of a single imputation.

The original full R code and a reproducible version illustrating the analysis using synthetic datasets are available in a GitHub repo (https://github.com/gasparrini/UKB-confounding).

Results

Descriptive analysis

The original dataset included 502,381 individuals. We excluded 4,363 (0.87%) participants due to missing data in the exposure history, with a final cohort of 498,018 people. The participants were followed up for an average of 11.19 years, with a total of 5,575,253 person-years. During the follow-up, we observed 37,878 deaths for non-accidental causes.

The descriptive analysis of spatial and temporal patterns in $PM_{2.5}$ exposure and mortality rates can be useful to illustrate related confounding mechanisms and adjustment strategies. Figure 2 shows the time series of the distribution of annual $PM_{2.5}$ exposure and mortality rates over the timescales of calendar year and age. The contrasting trends in exposure and mortality risks indicate a threat of temporal confounding, and the need for an adequate control in the regression model. Figure 3 shows the average distribution of $PM_{2.5}$ against mortality rates calculated within assessment centre (see Figure S1 in the appendix for their geographical locations). There is a south-to-north pattern, with higher $PM_{2.5}$ levels and lower mortality in centres belonging to the London area (Hounslow, Barts and Croydon) and partly in the South of England. The negative correlation makes the analysis prone to spatial confounding, which must be appropriately controlled for if not captured by area-level predictors.

Finally, Table 1 shows the distribution of average $PM_{2.5}$ exposure and mortality rates in categories of contextual and individual-level variables, indicating noticeable correlations. Specifically, higher area-level deprivation corresponds to higher mortality rates and strongly increasing exposure levels, while less urbanized areas are characterized by lower mortality rates and reduced air pollution. Similarly, residential greenness is associated with decreasing mortality and $PM_{2.5}$ concentrations. Individual-level variables show interesting patterns. For instance, average $PM_{2.5}$ exposure and mortality change across categories of both socio-economic and lifestyle factors, although part of these differences can be explained by age and

location. While variations in exposure are shallow, the striking differentials in mortality rates can still induce mild changes in the exposure-response associations.

Analysis of the directed acyclic graph

The assessment of the DAG in Figure 1 indicates that all the measurable risk factors listed above must be included in the regression model to adjust for confounding. We highlight here some interesting aspects, however acknowledging that these considerations depend on the strong, untested, and (to some extent) subjective assumptions about the causal paths. First, control for assessment centre is important to limit spatial confounding, occurring primarily through the path involving unobserved area-level characteristics linked with differential outdoor $PM_{2.5}$ levels. Second, a lack of control for individual-level factors, represented by both socio-economic conditions and lifestyle factors, can lead to confounding. Lifestyle characteristics can confound the association of outdoor $PM_{2.5}$ and mortality via multiple paths involving unobserved area-level characteristics and variation in personal $PM_{2.5}$ exposure. The mechanism involving individual-level socio-economic factors is more complex, as they do not confound directly, but need to be adjusted to close a backdoor path opened when controlling for lifestyle variables acting as colliders via unobserved area-level characteristics.

Regression analysis and control for confounding

The association between long-term exposure to $PM_{2.5}$ exposure and non-external mortality was first estimated with a fully adjusted model. The specification of this model provides control for the different types of confounding mechanisms mentioned above. First of all, the model directly adjusts for several personal characteristics and behaviours. Additionally, the inclusion assessment centre and year of birth as stratifying variables, together with sex, ensures a strong control for spatial and temporal patterns for $PM_{2.5}$ and death trends. Finally, additional control for contextual variables is provided by area-level deprivation, neighbour greenspace, and urban/rural classification. This main model (Model 1) reports an HR of 1.25 (95%CI: 1.06,1.49) for a 10 μg/m³ increase in the exposure, as shown in Table 2. The various confounding mechanisms were examined and quantified by fitting models with alternative specifications, with the corresponding estimates reported in the rest of the table.

The top part of Table 2 concerns temporal confounding, presenting models where calendar trends and age are adjusted in different ways. Specifically, age does not seem to act as a confounder, as the estimated HR is very close to the main model when it is controlled directly through spline terms (Model 2), or not controlled at all (Model 3). The situation is different for calendar time: when using age as the time axis, calendar trends need a strict adjustment through stratification by year (Model 4) or splines (Model 5). More importantly, the absence of control for calendar trends (Model 6) results in underestimation compared to the main model, with an HR of 1.18 (1.06-1.31).

The following models inform about different spatial confounding mechanisms, with more pronounced variations for different confounder adjustments. Specifically, the removal of control for assessment centre (Model 7), used in the main model as a stratifying variable, reveals a large confounding effect, with the risk reversing to an HR of 0.82 (0.76-0.87). This is consistent with the strong geographical correlation between baseline mortality and $PM_{2.5}$ exposure shown in Figure 2. Also notable is the much narrower confidence intervals compared to the main model, related to the wider exposure contrast obtained when removing assessment centre as a stratifying variable. The sensitivity analyses conducted for the inclusion of assessment centre (Table S1) show consistency among most methods (strata, indicator and gamma random effects) except when normally distributed random effects are used.

Other spatial mechanisms can be related to variations in the distribution of measurable contextual risk factors, which can be assessed by removing each of them in turn from the model (Models 8-10). The comparison demonstrates a strong confounding effect of area-level deprivation, with the HR increasing to 1.51 (1.28-1.80) when excluding it from the model, while the control for greenspace and urban-rural produces minimal changes. Finally, the last two models examine the role of individual-level risk factors, reporting estimates where socio-economic variables (Model 11) or other personal characteristics (Model 12) are removed, respectively. While the former does not seem to exert any confounding effect, the lack of control for physical and lifestyle factors seems to lead to a noticeable overestimation, with an HR of 1.34 (1.12-1.59). An additional analysis where we evaluated the contribution from each of these factors separately indicated that most of this confounding effect seemed to be exerted by smoking behaviours, in particular the current status (Table S3).

Discussion

In this contribution, we illustrate and discuss confounding mechanisms in epidemiological studies on long-term risks associated with environmental factors, with a specific example on the association between $PM_{2.5}$ and mortality in the UK Biobank. We perform a comprehensive assessment including descriptive analyses, theoretical considerations based on DAGs, and empirical results from regression models. The results are consistent and reveal various confounding patterns, offering insights as well as advice on specific adjustment strategies to be implemented in this context.

The most striking result is the strong spatial confounding linked with the lack of control for assessment centre, which represents a proxy for the residential area of the subject. The rationale of stratifying for assessment centres in the Cox model is to limit the comparison within each risk set to people living in the same area, thus reducing differences in unobserved contextual risk factors. Failing to adjust for it leads to a strong confounding effect, with the HR reversing from 1.25 to 0.82, due to the strong correlation with both $PM_{2.5}$ concentrations and baseline mortality rates. It should be noted, however, that there is a trade-off between an aggressive spatial stratification and the reduction in exposure variation within risk sets that leads to reduced statistical power, as shown in the results. Some published analyses introduced spatial control, particularly through area-level random effects[15,16], but others have not[17–21]. Our result showing differential estimates depending on the distribution of random effects (Table S2) is interesting, and it deserves further scrutiny in future analyses. In any case, these results indicate that more attention needs to be devoted to controlling for spatial confounding in cohort studies on environmental risks. In this analysis, the negative direction of the association stems from the opposite geographical patterns of air pollution and death rates shown in Figure 3, consistent with the “Glasgow effect” previously reported in the literature, which represents an excess mortality in the Scottish population due to historical socio-economic circumstances[22].

Another interesting aspect is the issue of temporal confounding, requiring considerations of the choice of the time axis and the direct control for trends across other temporal dimensions. Our association estimates were invariant to the inclusion of age as a predictor, regardless of its specification, but highly sensitive to the exclusion of calendar year, leading to an overestimation of the health risks when the latter was excluded. These results can be explained by multiple aspects, including the use of time-varying exposure measures, the opposing trends in $PM_{2.5}$ and mortality over the study period, and the UKB being a closed (fixed) cohort, with the recruitment occurring in a short period and no replacement. These features motivated our choice regarding the selection of the time axis and the control of other temporal trends. While such a choice can change depending on the study setting and design,

as well as on the temporal resolution of the exposure or other covariates, we emphasise the importance of appropriately accounting for temporal confounding in the analysis of long-term effects of environmental stressors.

One important topic addressed in this contribution is the confounding mechanisms related to individual-level factors on health risks associated with air pollution. The question is motivated by the lack of individual information in administrative health databases often used to perform large population-based cohort analyses[23,24], other than basic socio-economic indices (e.g., education and income) and ethnicity. In particular, the role of lifestyle characteristics is currently debated, with a line of thought asserting that these factors should not be controlled for, as they are not correlated with area-level pollution levels[25]. Both theoretical arguments and empirical results presented here put such an argument into question. First, the DAG in Figure 1 demonstrates that lifestyle factors can confound through multiple pathways, for instance, because of spatial correlation with pollution levels or via their role in altering personal exposure, both induced by the presence of unobserved area-level characteristics. Second, the regression analysis indicates a noticeable increase in HR when removing lifestyle variables, leading to an overestimation of the risk. This finding aligns with previous studies that found strong air pollution effects when accounting for individual-level covariates[26–28]. In particular, our associations are analogous to those estimated with a large Canadian cohort[26], in which the addition of behavioural covariates decreased the HR for mortality from 1.36 to 1.26 after adjustment for individual- and area-level factors. This result is also consistent with the descriptive analysis, where even very small variations in exposure distributions across categories for lifestyle factors can lead to confounding due to the effects of these variables, as shown in Table 1. Indeed, variables such as smoking, alcohol consumption, and obesity represent by far the strongest risk factors for mortality.

Some limitations must be acknowledged. First, the results presented here are specific to the UK Biobank. Nonetheless, while different patterns can be found in other cohorts, general considerations about confounding mechanisms would still be relevant. Second, the UKB database presents some limitations, in particular the fact that most of the variables are collected only at recruitment and that the residential locations of the subjects are not made available to the users, requiring the use of assessment centre as a proxy for geographical areas. Both of these issues surely introduce errors and imprecision, although again, this fact does not affect the general considerations made here about confounding. Finally, it must be stressed that this contribution aims to identify and discuss various confounding patterns in cohort studies on the health risks of environmental stressors, albeit without making use of and discussing complex causal inference methodologies to address them.

In conclusion, this case-study illustration reveals various confounding effects in cohort studies on long-term risks of environmental exposures, linked with spatial and temporal data structures in addition to direct contributions from contextual and individual-level factors. The results show consistent patterns between descriptive analyses, theoretical arguments from DAGs, and empirical results from regression models. The findings offer insights into approaches to control for spatial differences and temporal trends in exposure and risks, as well as the need to adjust for a set of potential confounders, including lifestyle variables.

Funding

This work was supported by the Nagasaki University "Doctoral Program for World-leading Innovative and Smart Education" for Global Health (WISE), KENKYU SHIDO KEIHI ("the Research Grant") and KYOIKU KENKYU SHIEN KEIHI ("the Stipend"), and by the Medical Research Council UK Grant ID: MR/Y003330/1).

Acknowledgements

This research has been conducted using the UK Biobank Resource under application number 56431. This work was part of a PhD project supported by the Nagasaki University WISE Programme.

Data and reproducibility

The original full R code used to perform the analysis is available in a GitHub repo (https://github.com/gasparrini/UKB-confounding) and it can be used alongside the synthetic datasets that resemble the original UK Biobank data. The actual database of the UK Biobank can be accessed via an application through the UK Biobank online Access Management System (AMS) (https://www.ukbiobank.ac.uk/). The $PM_{2.5}$ exposure dataset was expressly linked in this analysis and will be fully attached to the main database and released in the UK Biobank repository.

Conflict of interests' statement

The authors declare that they have no competing interests.

References

1. Pope, C. A., 3rd, Coleman, N., Pond, Z. A. & Burnett, R. T. Fine particulate air pollution and human mortality: 25+ years of cohort studies. *Environ Res* **183**, 108924 (2020).

2. Chen, J. & Hoek, G. Long-term exposure to PM and all-cause and cause-specific mortality: A systematic review and meta-analysis. *Environment International* **143**, 105974 (2020).

3. Dominici, F., Sheppard, L. & Clyde, M. Health Effects of Air Pollution: A Statistical Review. *International Statistical Review* **71**, 243–276 (2003).

4. Sheppard, L. *et al.* Confounding and exposure measurement error in air pollution epidemiology. *Air Qual Atmos Health* **5**, 203–216 (2012).

5. Papadogeorgou, G., Kioumourtzoglou, M. A., Braun, D. & Zanobetti, A. Low Levels of Air Pollution and Health: Effect Estimates, Methodological Challenges, and Future Directions. *Current Environmental Health Reports* **6**, 105–115 (2019).

6. Li, C. *et al.* Reversal of trends in global fine particulate matter air pollution. *Nat Commun* **14**, 5349 (2023).

7. Griffin, B. A., Anderson, G. L., Shih, R. A. & Whitsel, E. A. Use of alternative time scales in Cox proportional hazard models: implications for time-varying environmental exposures. *Statistics in Medicine* **31**, 3320–3327 (2012).

8. Bind, M.-A. Causal Modeling in Environmental Health. *Annual Review of Public Health* **40**, 23–43 (2019).

9. Bauwelinck, M. *et al.* Variability in the association between long-term exposure to ambient air pollution and mortality by exposure assessment method and covariate adjustment: A census-based country-wide cohort study. *Science of The Total Environment* **804**, 150091 (2022).

10. Samoli, E. *et al.* Modeling multi-level survival data in multi-center epidemiological cohort studies: Applications from the ELAPSE project. *Environment International* **147**, 106371 (2021).

11. Sudlow, C. *et al.* UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS medicine* **12**, e1001779 (2015).

12. de la Cruz Libardi, A. *et al.* High resolution mapping of nitrogen dioxide and particulate matter in Great Britain (2003–2021) with multi-stage data reconstruction and ensemble machine learning methods. *Atmospheric Pollution Research* **15**, 102284 (2024).

13. Vanoli, J. *et al.* Reconstructing individual-level exposures in cohort analyses of environmental risks: an example with the UK Biobank. *J Expo Sci Environ Epidemiol* 1–6 (2024) doi:10.1038/s41370-023-00635-w.

14. Jacopo Vanoli *et al.* Long-term associations between time-varying exposure to ambient PM2.5 and mortality: an analysis of the UK Biobank. *Epidemiology* **In press**,.

15. Carey, I. M. *et al.* Mortality Associations with Long-Term Exposure to Outdoor Air Pollution in a National English Cohort. *Am J Respir Crit Care Med* **187**, 1226–1233 (2013).

16. Beelen, R. *et al.* Long-term Exposure to Air Pollution and Cardiovascular Mortality: An Analysis of 22 European Cohorts. *Epidemiology* **25**, 368 (2014).

17. Strak, M. *et al.* Long term exposure to low level air pollution and mortality in eight European cohorts within the ELAPSE project: pooled analysis. *BMJ* **374**, n1904 (2021).

18. Stafoggia, M. *et al.* Long-term exposure to low ambient air pollution concentrations and mortality among 28 million people: results from seven large European cohorts within the ELAPSE project. *Lancet Planet Health* **6**, e9–e18 (2022).

19. Jiang, Z. *et al.* Co-exposure to multiple air pollutants, genetic susceptibility, and the risk of myocardial infarction onset: a cohort analysis of the UK Biobank participants. *European Journal of Preventive Cardiology* zwad384 (2023) doi:10.1093/eurjpc/zwad384.

20. Cai, Y. *et al.* Road traffic noise, air pollution and incident cardiovascular disease: A joint analysis of the HUNT, EPIC-Oxford and UK Biobank cohorts. *Environment International* **114**, 191–201 (2018).

21. Doiron, D. *et al.* Air pollution, lung function and COPD: results from the population-based UK Biobank study. *European Respiratory Journal* **54**, (2019).

22. Walsh, D., McCartney, G., Collins, C., Taulbut, M. & Batty, G. D. History, politics and vulnerability: explaining excess mortality in Scotland and Glasgow. *Public Health* **151**, 1–12 (2017).

23. Di, Q. *et al.* Air Pollution and Mortality in the Medicare Population. *New England Journal of Medicine* **376**, 2513–2522 (2017).

24. Pappin, A. J. *et al.* Examining the Shape of the Association between Low Levels of Fine Particulate Matter and Mortality across Three Cycles of the Canadian Census Health and Environment Cohort. *Environmental Health Perspectives* **127**, 107008 (2019).

25. Weisskopf, M. G. & Webster, T. F. Trade-offs of Personal Versus More Proxy Exposure Measures in Environmental Epidemiology. *Epidemiology* **28**, 635 (2017).

26. Pinault, L. *et al.* Risk estimates of mortality attributed to low concentrations of ambient fine particulate matter in the Canadian community health survey cohort. *Environmental Health* **15**, 18 (2016).

27. Raaschou-Nielsen, O. *et al.* Long-term exposure to air pollution and mortality in the Danish population a nationwide study. *eClinicalMedicine* **28**, (2020).

28. Bentayeb, M. *et al.* Association between long-term exposure to air pollution and mortality in France: A 25-year follow-up study. *Environment International* **85**, 5–14 (2015).

Tables

Table 1. Distributions of subjects, fine particulate matter (PM2.5) and mortality across categories of contextual and individual-level variables in the UK Biobank cohort (see Table S1 in the appendix for further details).

Variable	Category	Count (%)	PM2.5 average (SD)	Death rate (x 100,000)
Gender	Female	271,611 (54.5%)	10.79 (1.85)	518.13
	Male	226,407 (45.5%)	10.80 (1.83)	876.74

	Missing (%)	0 (0.0%)	NA	-
Townsend deprivation index (TDI)	1th quintile	196,400 (39.4%)	10.40 (1.56)	594.31
	2nd quintile	180,916 (36.3%)	10.79 (1.75)	651.38
	3rd quintile	83,745 (16.8%)	11.31 (2.08)	792.83
	4th quintile	33,752 (6.8%)	11.84 (2.38)	1,011.62
	5th quintile	3,205 (0.6%)	11.82 (2.81)	1,285.01
	Missing (%)	0 (0.0%)	-	-
Urban-rural classification	Urban	424,129 (85.2%)	10.98 (1.83)	689.37
	Town/fringe	37,611 (7.6%)	9.74 (1.53)	658.8
	Village/Rural	36,278 (7.3%)	9.70 (1.45)	584.34
	Missing (%)	0 (0.0%)	-	-
Greenspace	1th quintile	82,036 (16.5%)	12.19 (2.07)	636.42
	2nd quintile	168,841 (33.9%)	11.01 (1.74)	713.31
	3rd quintile	127,046 (25.5%)	10.46 (1.52)	700.45
	4th quintile	78,132 (15.7%)	10.04 (1.47)	655.79
	5th quintile	41,963 (8.4%)	9.64 (1.44)	607.59
	Missing (%)	0 (0.0%)	-	-
Ethnic background	White	470,847 (94.5%)	10.71 (1.81)	692.56
	Other	27,171 (5.5%)	12.27 (1.81)	447.45
	Missing (%)	0 (0.0%)	-	-
Education	Low	85,753 (17.2%)	10.69 (1.76)	1,245.11
	Professional	59,242 (11.9%)	10.66 (1.77)	789.52
	High school	190,183 (38.2%)	10.74 (1.75)	565.25
	College	162,840 (32.7%)	10.96 (2.00)	485.17
	Missing (%)	0 (0.0%)	-	-
Income	<18,000	117,126 (23.5%)	10.82 (1.81)	1,156.19
	18,000-30,999	124,125 (24.9%)	10.72 (1.78)	739.34
	31,000-51,999	127,779 (25.7%)	10.74 (1.81)	494.24
	52,000-100,000	101,486 (20.4%)	10.82 (1.89)	390.28
	>100,000	27,502 (5.5%)	11.21 (2.13)	389.34
	Missing (%)	0 (0.0%)	-	-
Employment	Employed	288,733 (58.0%)	10.84 (1.86)	369.85
	Retired	166,525 (33.4%)	10.67 (1.78)	1,176.45
	Other	42,760 (8.6%)	10.99 (1.94)	908.99
	Missing (%)	0 (0.0%)	-	-
Smoking status	Never	273,466 (54.9%)	10.75 (1.83)	480.1
	Previous	172,100 (34.6%)	10.82 (1.84)	831.45
	Current	52,452 (10.5%)	10.98 (1.91)	1,253.93
	Missing (%)	0 (0.0%)	-	-
Smoking intensity	0	275,782 (55.4%)	10.75 (1.83)	484.27

	<=10	60,933 (12.2%)	10.90 (1.86)	510.44
	Oct-30	104,544 (21.0%)	10.85 (1.85)	793.66
	30-60	48,488 (9.7%)	10.81 (1.85)	1,518.47
	>60	8,271 (1.7%)	10.85 (1.87)	2,537.80
	Missing (%)	0 (0.0%)	-	-
Alcohol drinking status	Never	40,138 (8.1%)	11.05 (1.91)	986.13
	Occasionally	57,554 (11.6%)	10.93 (1.86)	805.66
	1-3 a month	55,588 (11.2%)	10.76 (1.81)	587.34
	1-2 a week	128,677 (25.8%)	10.67 (1.79)	604.3
	3-4 a week	114,975 (23.1%)	10.70 (1.82)	564.93
	Daily or almost	101,086 (20.3%)	10.91 (1.88)	768.01
	Missing (%)	0 (0.0%)	-	-
Waist-to-hip ratio	low	253,546 (50.9%)	10.77 (1.87)	528.81
	medium	129,153 (25.9%)	10.78 (1.82)	749.01
	high	115,319 (23.2%)	10.87 (1.82)	937.97
	Missing (%)	0 (0.0%)	-	-
Physical activity (IPAQ score)	Low	92,953 (18.7%)	10.78 (1.80)	793.6
	Moderate	193,796 (38.9%)	10.81 (1.88)	662.65
	High	211,269 (42.4%)	10.78 (1.83)	644.98
	Missing (%)	0 (0.0%)	-	-
Living alone	No	404,368 (81.2%)	10.74 (1.81)	607.5
	Yes	93,650 (18.8%)	11.02 (1.97)	998.36
	Missing (%)	0 (0.0%)	-	-

471

Table 2. Hazard ratio (HR) associated with an increase of 10 μg/m³ in $PM_{2.5}$ in the UK Biobank cohort, estimated from models with different adjustment strategies, with 95% confidence intervals (CI).

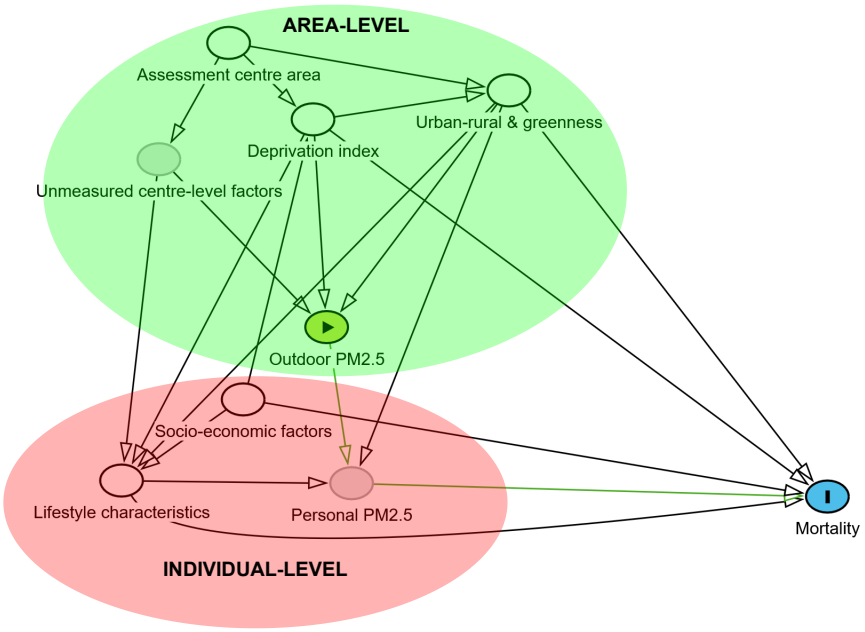
Main adjustment	**Temporal adjustment**		**Model**	**HR (95% CI)**
	Time axis	Additional term		
Fully-adjusted	Calendar	Strata of age (months)	Model 1	1.25 (1.06-1.49)
	Calendar	Splines of age	Model 2	1.23 (1.04-1.47)
	Calendar	-	Model 3	1.25 (1.06-1.49)
	Age	Strata of calendar (months)	Model 4	1.24 (1.04-1.47)
	Age	Splines of calendar time	Model 5	1.27 (1.08-1.49)
	Age	-	Model 6	1.18 (1.06-1.31)
No assessment centre	Calendar	Strata of age (months)	Model 7	0.82 (0.76-0.87)
No deprivation index	Calendar	Strata of age (months)	Model 8	1.51 (1.28-1.80)
No urban-rural classification	Calendar	Strata of age (months)	Model 9	1.27 (1.07-1.51)
No greenspace	Calendar	Strata of age (months)	Model 10	1.24 (1.06-1.46)
No individual-level SES factors	Calendar	Strata of age (months)	Model 11	1.25 (1.05-1.48)
No individual-level lifestyle factors	Calendar	Strata of age (months)	Model 12	1.34 (1.12-1.59)

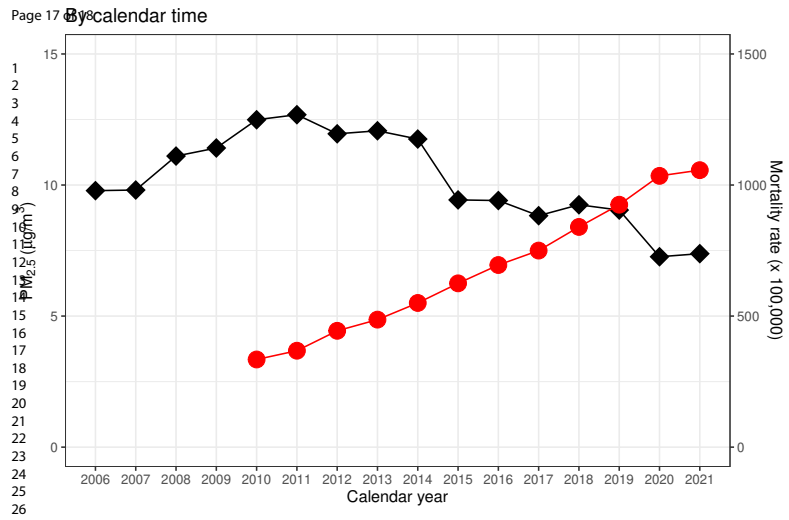
Figures

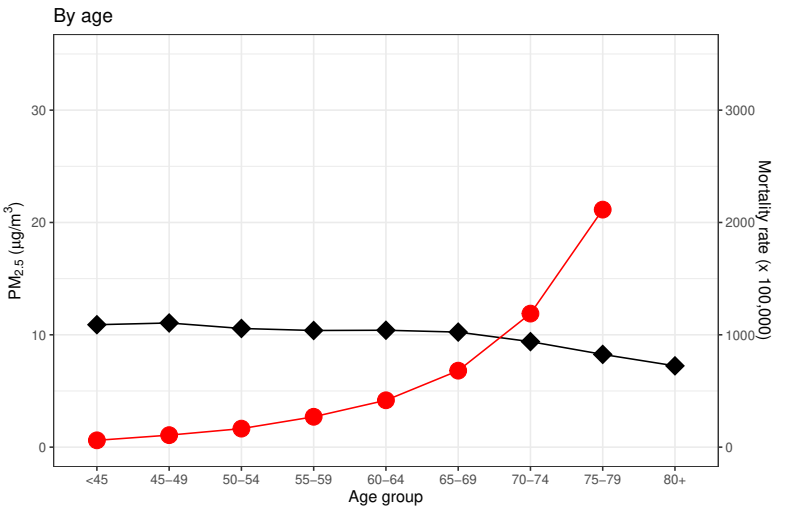
Figure 1. Directed acyclic graph (DAG) representing the assumed causal paths between exposure to fine particulate matter ($PM_{2.5}$) and mortality in the UK Biobank cohort.

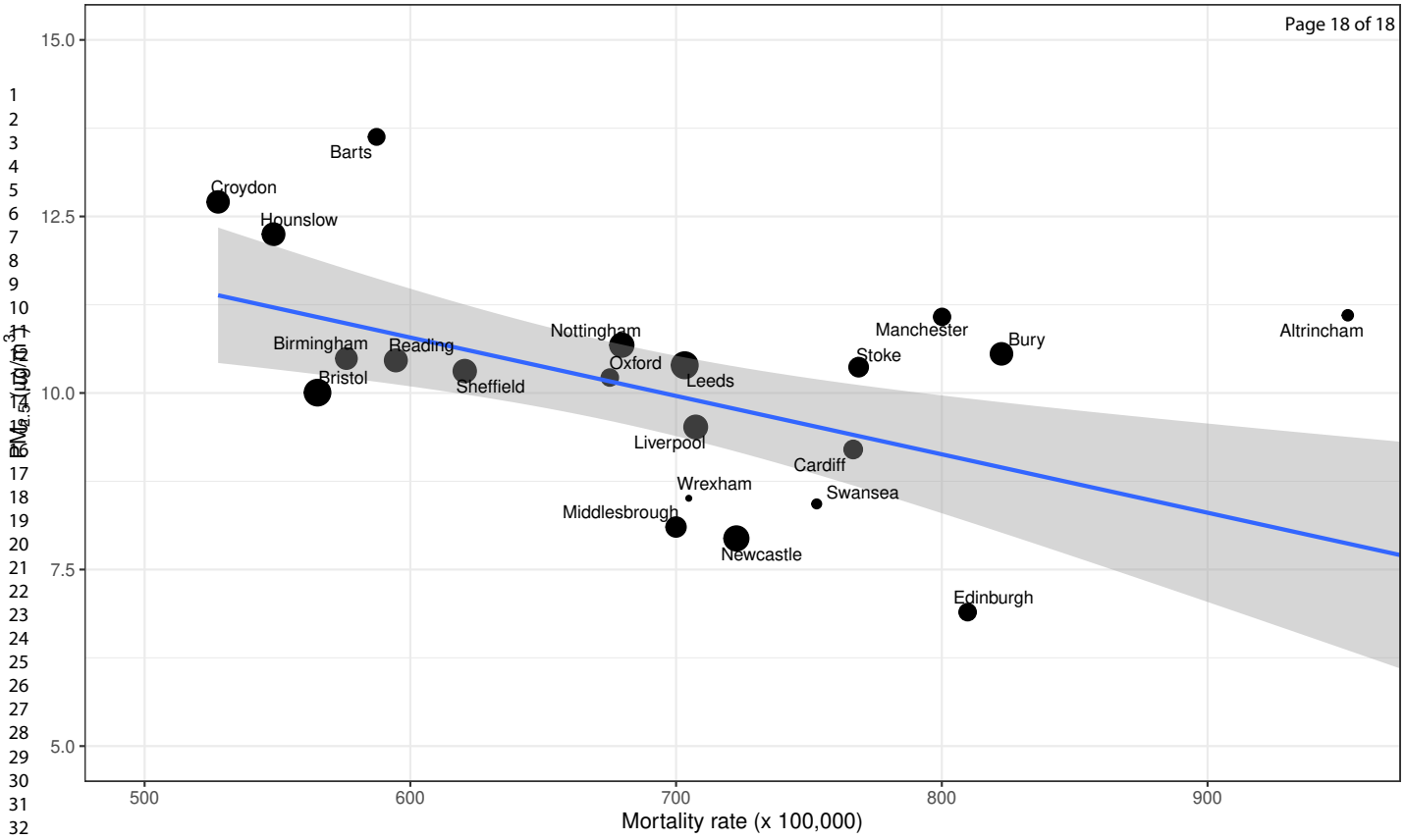
Figure 2. Trends of annual average exposure to fine particulate matter ($PM_{2.5}$, in $\mu g/m^3$, black squared line) and mortality rates (per 100,000 person-years) by calendar time (left, red dotted line) and age (right, red dotted line) in the UK Biobank cohort.

Figure 3. Scatterplot of average exposure to fine particulate matter ($PM_{2.5}$, in $\mu g/m^3$) and mortality rates (per 100,000 person-years) by assessment centre in the UK Biobank cohort.


AREA-LEVEL
Assessment centre area
Urban-rural & greenness
Deprivation index
Unmeasured centre-level factors
Outdoor PM2.5
Socio-economic factors
Lifestyle characteristics
Personal PM2.5
Mortality
INDIVIDUAL-LEVEL


By calendar time
15
10
5
0
1500
1000
500
0
2006
2007
2008
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019
2020
2021
Calendar year
Mortality rate (x 100,000)


By age
30
20
10
0
3000
2000
1000
0
<45
45–49
50–54
55–59
60–64
65–69
70–74
75–79
80+
Age group
Mortality rate (x 100,000)


15.0
12.5
10.0
7.5
5.0
500
600
700
800
900
Mortality rate (x 100,000)
Barts
Croydon
Hounslow
Birmingham
Reading
Bristol
Sheffield
Nottingham
Oxford
Leeds
Liverpool
Wrexham
Middlesbrough
Newcastle
Cardiff
Swansea
Stoke
Manchester
Bury
Altrincham
Edinburgh

Chapter 13

Final comments

In this concluding chapter, I offer final remarks on my doctoral research concerning the evaluation of the long-term impact of air pollution within the UK Biobank cohort. First, I will present the main outputs of this PhD and their relationship with the PhD's objectives. In the second section, I will present the contribution to the scientific literature of the output of this PhD project. Finally, I will examine possible future developments from computational, theoretical, and applied perspectives. A final discussion concludes the thesis.

13.1 Main outputs

The main findings from the studies conducted and presented in the previous chapters are summarised here.

In Chapter 9, I presented a general framework for linking environmental exposure data with cohort databases. The procedure consists of three subsequent steps focusing on the description of exposure masking methods and

possible uses of the output data in epidemiological studies. This process was illustrated by connecting daily $PM_{2.5}$ predictions with UK Biobank participants' residential addresses. The linkage described fulfilled the Objective I of the PhD and it was practically applied to perform two epidemiological analyses (Objective II and III) on health outcomes with the UK Biobank cohort.

In Chapter 10 and 11 I presented two long-term studies on the effect of $PM_{2.5}$ on the risk of early death and cardiovascular inpatient hospital admissions, respectively. The studies were conducted with a follow-up of up to 13 years per subject with an overall exposure history of 19 years. Annual time-varying averages of $PM_{2.5}$ were linked to the UK Biobank participants' residential locations. Then, Cox PH models were applied to estimate the exposure-response functions for each outcome in separate models.

In the mortality analysis (Chapter 10) an increase of 10 $\mu g/m^3$ in 8-year averages of $PM_{2.5}$ exposure was associated with all-cause [HR= 1.28 (95% CI: 1.07-1.53)], non-accidental [1.24 (1.03-1.51)], respiratory [2.07 (1.04-4.11)] and cancer mortality [1.66 (0.86-3.19)] in separate models. No association was found for cardiovascular causes for that time frame [0.89 (0.60-1.31)]. Moreover, the application of different time frames and distributed lag exposure-response risk functions suggested diverse patterns of association among the outcomes, but the statistical uncertainty prevented making clear conclusions. However, the elevated risk for respiratory mortality appears to be primarily linked to the last three years of exposure and the cardiovascular mortality risk in the previous year. Finally, this methodological development (using DLMs in long-term studies) can also be applied in future UKB studies and with other cohorts.

In the analysis of hospital admissions (Chapter 11), positive significant associations were detected between the long-term average of $PM_{2.5}$ and multiple cardiovascular outcomes, including MI NSTEMI (non-ST-elevation) intracerebral stroke, heart failure and cardiac arrest. The effects were mostly consistent across exposure windows, except for MI NSTEMI, for which the associations were only positively significant with the 1-year average. For several outcomes, non-linear exposure-response functions showed linear associations and steep risk increase at lower levels of exposure, below 10 $\mu g/m^3$, suggesting the strong effects to be below international and national standard limits.

From the covariates adjustment perspective, both the mortality and hospital admissions analyses were sensitive to including individual- and area-level variables in the survival models. Compared to the basic model, adjusted for assessment centre, sex and age, the most significant change in the hazard ratios occurred after control for area deprivation index, albeit also the inclusion of greenspace, population density (urban/rural) and individual characteristics and behaviours affected the health effect estimates.

In the last contribution (Chapter 12), I investigated confounding mechanisms in the relationship between long-term $PM_{2.5}$ and non-accidental mortality. The analysis was carried out using both statistical descriptives and survival regression models, which were applied using the design of the above-described studies without the investigation of the temporal windows and lagged effects. The analyses showed an important confounding impact across time (calendar years and participants' age) and geographical areas. On the temporal scale, mortality rates tend to increase both in time and naturally with ageing, albeit $PM_{2.5}$ trends display decreasing patterns. On the spatial dimension, the

exclusion of the assessment centres from the models reverses the association between $PM_{2.5}$ and mortality, suggesting potential strong confounding effects when knowledge of large geographical areas is unaccounted for. Finally, results also showed the associations changed moderately but significantly when individual characteristics and behaviours were excluded from the models. Theoretical justifications for this potential confounding mechanisms were drawn and a directed acyclic graph (DAG) was defined.

13.2 Contributions

This section highlights this thesis's contributions to the scientific literature and its implications for public health.

13.2.1 Exposure linkage framework

The linkage procedure applied is similar to previous studies that linked environmental exposure measures to cohort data. However, the linkage performed for this PhD is the first attempt to describe a linkage process with highly resolved spatiotemporal data illustrating the detail in every step. This has been helping and will continue to contribute to enabling a wider audience to understand how ambient exposure linkages are structured and function. Moreover, this work treats the problem of privacy-protecting methods in environmental epidemiology studies and justifies the choice of bilinear interpolation as a potentially efficient and effective method for the study purposes. While this linkage procedure was used with the UK Biobank cohort, it can be virtually applied to any pair of environmental exposure-cohort databases that include information on the residential histories. This will simplify the

work of future or less experienced researchers searching for established linkage procedures to apply to their datasets. Finally, the exposure data will be available to the UKB research community, allowing investigators to conduct a wide range of epidemiological analyses involving short to long-term effects analyses.

13.2.2 Air pollution effects on mortality in the context of the literature

A rich literature exists describing the effects of long-term $PM_{2.5}$on mortality. However, among all the investigations, only a few included resolved spatiotemporal exposure information in their design, and most were conducted in North America. Therefore, this is one of the few studies using highly resolved time-varying exposure in Europe, and the first study of this kind was performed using a UK-based cohort.

Because the heterogeneity among previous studies did not favour the extrapolation of health risks outside the area and populations investigated, the current study is a relevant addition as it corroborates the scarce literature of UK-based studies on the topic.

This study provides updated associations using state-of-the-art methods. Compared to the most comprehensive literature reviews to date[27, 28], this study shows stronger associations between 8-year $PM_{2.5}$ averages and premature all-cause mortality against meta-analytical estimates of 1.09 (1.07, 1.11). Similarly, remarkably strong effects were found for respiratory and lung cancer mortality, even though the results displayed high imprecision.

In contrast with the previous studies[27, 112], there was a null effect for cardiovascular mortality.

Our results for cause-specific deaths are in line with the only previous individual-level study conducted on a UK-based cohort[44], which showed strong effects for respiratory mortality but no significant associations with lung cancer and cardiovascular deaths after adjustments for area-level SES. A reason for this could be the relatively short follow-up and young age of the cohort. Another reason could be attributable to the complex relationship between ambient air pollution and SES variables that prevent drawing firm conclusions. Further studies should investigate this weak association. Unexpectedly, compared to a previous UK Biobank study that analysed the risk of mortality due to $PM_{2.5}$ using 15-year-old exposure predictions from ESCAPE, the estimates of the current study were similar for all-cause mortality and lower for cardiovascular mortality.

This was the first study to assess the temporal decomposition of mortality risk using DLM. These analyses could be important to identify the most relevant windows through air pollution exerts its most detrimental effects across the follow-up.

13.2.3 Air pollution effects on cardiovascular hospital admissions in the context of the literature

Regarding major adverse cardiovascular events (MACE), across all the time windows, UK resident adults living in areas with a 5 $\mu g/m^3$ higher exposure experienced a 12% to 15% elevated risk of MACE-related hospitalisations compared to those in less exposed areas. These results partially align with previous studies[113–115], even though their comparison is difficult due to

varying MACE definitions. The unclear results for acute myocardial infarction align with previous meta-analyses[26, 116]. Of note, this is the first long-term study investigating MI subtypes (STEMI and NSTEMI) and found that NSTEMI was associated with air pollution. At the same time, the association was less clearly defined for STEMI. This highlights the importance of focusing on specific health outcomes instead of investigating general ICD codes.

Previous studies[115, 117] using time-varying exposure have assigned the exposure average of one year prior to the event. In this study, exposure windows of varying lengths were assigned to the participants and analysed in separate models. In general, variations of the exposure windows only lead to mild changes in the associations. However, in few cases the changes were more pronounced. In particular, some cardiovascular outcomes, such as intracerebral stroke and non-ST-elevation myocardial infarction (NSTEMI) appear to be more sensitive to shorter (1-year) exposure windows compared to longer ones (5-year).

The non-linear concentration-response function and subset analysis indicate steep risks at concentrations across the exposure distribution, with no evidence of a threshold at the lowest values. This result underscores that despite recent decreases in air pollution levels, adverse effects persist even at very low concentrations. This finding adds to the growing body of literature in the rest of the Western countries, highlighting the importance of addressing air pollution concerns at both high and low exposure levels.

13.2.4 Effects at low levels

In the last 30 years, a lot of research has been attempting to investigate potential correlations between ambient long-term air pollution and increased risk of adverse health events. Milestone studies from the early 90s[24, 37] to 2000[36, 109] suggested that air pollution, mainly represented by PM_{10} (particulate matter with diameter <10 $\mu g/m^3$), $PM_{2.5}$ (fine particulate matter), SO_2 and Black Smoke (BS), contributed to excess all-cause and cause-specific mortality (cardiovascular and respiratory) in both North American and European urban areas. These findings motivated international and national health and environmental organisations to establish standardised threshold limits for air pollution. For instance, the WHO in 2005 set the annual limit for $PM_{2.5}$ at 10 $\mu g/m^3$. These historical results have been subsequently strengthened and validated by an updated and new series of publications[45–47] that found associations between exposure and mortality persisted at annual levels below the 10 $\mu g/m^3$ threshold, suggesting new threshold limits for ambient air pollutants across the whole world. Of particular note, important studies have shaped the literature by exploiting national representative cohorts[107], massive sample sizes[48, 49] and causal frameworks[118, 119], especially in North America, as well as implementing large multi-city and multi-cohort[39, 51] studies, in Europe. Numerous have also been attempts to shed light on the low levels (depending on the national standards) of air pollution and the potential different impact of components of PM. As a consequence, in 2021 the WHO re-evaluated the annual limit for $PM_{2.5}$ setting it at 5 $\mu g/m^3$.

The UK case In the rest of the Western countries, annual air pollution level limits have tightened; on the other hand, national UK limits have been

markedly more permissive. The 2010 Air Quality Standard regulations imposed the limits to the annual $PM_{2.5}$ concentrations to 20 $\mu g/m^3$, while the updated Environmental Targets for England set in 2023 require to limit the annual average for the pollutant to be below 10 $\mu g/m^3$ by 2040 [120]. Compared to those recommended by international health organisations such as the WHO, the less restrictive limits on air pollution set in the UK can be attributed to several factors, including a lack of studies focusing specifically on the effects of low-level air pollution in the UK context. Another important reason is the potential applicability of international studies on the impact of low levels of air pollution. While international research has consistently linked low levels of air pollution to adverse health effects, UK policymakers have sometimes questioned the direct relevance of these findings to the UK population, considering the differences in environmental, demographic, and socioeconomic contexts[121]. In this project, I corroborated the findings from international studies by carrying evidence of the effect of low levels from a new rich individual-level cohort, the UK Biobank, paired with high-resolution exposure measurements suited to address a broad range of research questions on various health outcomes.

13.3 Strengths and limitations

In this section I describe the limitations of the studies I conducted.

13.3.1 Strengths of the study

The linkage procedure along with the gridded exposure maps permitted the exposure assignment at relatively high spatial and temporal resolution.

The study used a large individual-level cohort, and it is one of the few studies to investigate the long-term effects of $PM_{2.5}$ on mortality in the UK.
Detailed healthcare history favoured investigating specific cardiovascular health outcomes instead of general definitions.
Thanks to the detailed individual exposure history, customised exposure summaries for the participants were possible, and potentially intricate temporal effects were investigated by defining time-varying averages across different exposure windows and distributed lag terms.
These studies allowed control for a large number of potential confounders. These include area- and individual-level SES variables, personal characteristics, and lifestyles.

13.3.2 Limitations of the study

The linkage performed has one main limitation: the exposure assigned to the study participants is limited to the residential address or, at best, to the work/study place. This does not account for the subject's daily mobility away from a fixed place. However, it would be relevant to consider movements, in particular commuting time and space, to adequately obtain the subject's daily exposure profile.

The primary limitation of the epidemiological studies is the limited generalizability of the results, attributable to the "healthy volunteer bias" in the UK Biobank cohort. Although representativeness is not essential for valid epidemiological inference, significant selection bias can lead to biased estimates if there are unmeasured risk factors that serve as colliders by being linked to both the exposure and the likelihood of selection. The lack of effects and the large imprecision found when investigating most of the outcomes can be

attributed to low power in the analyses, considering that the number of cases for cause-specific deaths varied between 2000 to 6000 events. However, previous investigations in which extensive control of potential confounders was applied showed similar imprecision ranges. Due to the low number of cases for the outcomes (both mortality and hospital admissions) in the study, I was also not able to conduct short-term effect analyses, which require very large sample sizes to detect associations. Additionally, within this work I used only $PM_{2.5}$ as exposure in the epidemiological analysis. However, on spatial scales smaller than 1x1 km there exist other relevant exposures that co-vary with PM and have been associated with health. For example, these include traffi-related pollutants, such as NO_2, noise, black carbon and UFPs. Regrettably, most of these exposures were not available for my analyses, preventing me from assessing their direct impact or their potential role as effect modifiers for $PM_{2.5}$. In the cases when they were available, their temporal and spatial resolutions were extremely limited and therefore I could not include them in models along with $PM_{2.5}$. Finally, confounders' data were only collected at baseline therefore, I could not account for changes in personal SES, characteristics and behaviours across the follow-up.

13.4 Final developments

During the PhD project, in drafting the publications included in Chapters 7-10, I have attempted to provide a methodological description of an exposure linkage framework, two in-depth examples of epidemiological analyses and an exploration of the confounding mechanisms in environmental studies of long-term associations in the UK Biobank. However, there remains

considerable potential in research pertaining to exposure assessment and the subsequent health effects associations utilising the UK Biobank, presenting a good opportunity for numerous new investigations.

Linkage with new exposures The paper presented in Chapter 9 provides a description and detailed illustration of a general process to link environmental exposures with individual cohort data. Even though analogous frameworks have been applied in the literature, the one I present is illustrated in three clearly defined steps and directly discusses some innovative aspects, such as exposure masking, by outlining a comparison of the different methodologies. On the other hand, a deeper analysis of the methods would require additional time, which was not available during this PhD and can be performed in separate undertakings. Future work could investigate how different interpolation methods impact the exposure profiles and the estimated health associations in epidemiological studies. Moreover, the framework applied in this project is customisable; in fact, each step can be adapted to suit the needs of different and novel exposure models. For example, this framework can be integrated with the output of cutting-edge exposure techniques. Novel assessment methods are surpassing models based on the use of fixed predictors' measurements by integrating them with data obtained from mobile monitoring devices, such as Google Street View cars[122]. With a view on the specific UK Biobank case, future developments can see the linkage of additional environmental exposures. First of all, spatiotemporal predictions, analogous to the one presented for $PM_{2.5}$, are currently in the process of being linked with the UKB database, including NO_2 and PM_{10}[123]. However, the linkage is not to be limited to air pollutants. For instance, an interesting

additional step would be to include greenness measurements to complement the current values in the UKB database based only on the year 2010. Additionally, more updated measurements of the Townsend Deprivation Index could be added to the values measured in 2010. Finally, the customizability of the UKB database and the flexible characteristics of the linkage framework allow for the inclusion of additional predictions from different exposure models. Comparing exposure models is essential when assessing health associations, as agreement or disagreement between models would reveal potential measurement error biases in either method[124].

Epidemiological investigations The two epidemiological studies carried out and presented in Chapter 10 and 11 include diverse types of analyses, spanning temporal decomposition of effects, exposure modelling comparisons, and non-linear associations. These investigations utilised state-of-the-art modelling techniques and addressed contemporary research questions in epidemiology. However, they still only exemplify a much wider range of possible analyses. Different designs could be applied to conduct epidemiological studies depending on the exposure time frame of interest. For instance, the availability of daily exposure levels linked at the individual level allows for conducting short-term analyses of various acute health events. There are no investigations of short- and medium-term health associations with this cohort. Moreover, the framework can be extended beyond event outcomes. The presence of follow-up assessments in the UKB makes it possible to investigate repeated measures of continuous and categorical outcomes. Previous studies have attempted to analyse repeated measures with this cohort but have always used landmark exposure data from 2010[96, 125]. Re-analyses of these

studies or new investigations using different outcomes would incredibly benefit from the existence of time-varying exposure data. Finally, the UK Biobank includes large amount of data on blood samples. The high resolution of this exposure data will allow to explore not only clinically evident outcomes, but also the pathophysiological mechanisms leading to them[91].

Analyses of confounding The paper on confounding effects in long-term environmental studies presented in chapter 12 explores the impact of diverse mechanisms in the relationship between air pollution and mortality. However, in my study, the associations presented are purely associational, and the work did not include the use of causal inference methods. These methods combine classic statistical modelling techniques with a mathematically coherent framework for the causal relationship between variables in a defined system. There are several examples of the application of causal methods in the context of environmental epidemiology[86, 126]. It would be of great importance to extend the current analysis of confounding with the UKB to new analyses in which different causal methods are compared against the traditional associational estimates[118]. Examples of causal methods include g-computation[127], inverse probability weighting[78] and propensity score adjustments[118] that could be applied to compute risk ratios on various health outcomes.

An additional development of this study from a causal perspective could involve extensions of the DAG to different spatial levels (centre, neighborhood level, near-address level). Such an extension would help to conduct studies investigating associations with highly spatially-varying pollutants (for instance, those traffic-related) as well as comparing how the effect of each pollutant

on health changes when assessed with different resolutions.

Another aspect related to confounding is the adjustment for lifestyle variables. As mentioned in the introductory sections, typically, such lifestyle variables are not available in large administrative cohorts, but methods of indirect adjustment have been developed for this situation. It has been investigated in only a few studies[55, 128], how direct and indirect adjustment results compare; therefore it would be interesting to use this well-phenotyped cohort to evaluate the indirect adjustment.

13.5 Conclusions

The work I presented in this thesis has focused on the definition of an exposure linkage as well as on a comprehensive assessment and exploration of the long-term effects of $PM_{2.5}$ on mortality. In particular, I focused on the temporal decomposition of the effects and disentangling confounding mechanisms. The exposure linkage framework defined here has multiple purposes; from a practical perspective, it may be useful to researchers working on the UKB cohort and other cohorts. Moreover, it is a valuable reference for graphically teaching students how to use GIS methods and how state-of-the-art environmental linkages are performed in the literature. In line with the previous literature, a relevant effect of $PM_{2.5}$ on mortality was observed. There were several significant associations between all-cause and cause-specific mortality as well as cardiovascular hospitalisations. This is one of the few UK-based study corroborating the literature produced in the rest of Europe and North America finding associations for several health outcomes. In general, the effect on mortality was not modified by the breadth of the exposure win-

dow considered. However, for mortality due to cardiovascular causes, the most relevant exposure frame to exert effects on health may be the closest to the event date. Similar results were obtained for cardiovascular hospital admissions, except for MI NSTEMI, which showed significant associations only with the shortest exposure window. This result may inform the possible etiological mechanisms linking air pollution and cardiovascular diseases compared to other health outcomes. Further research should investigate throughout mechanisms do $PM_{2.5}$ affect differential health risks. For most cardiovascular outcomes, the effect estimates exhibited sharp increases in cardiovascular risk at the lowest levels of exposure. This suggests, as previous research has done, that there is no safe threshold for exposure, highlighting the importance of defining more stringent standards in the UK as in other Western countries. It could also entail the need for enhanced monitoring of air quality and public awareness campaigns about the risks associated with low-level exposures. Finally, it appears imperative to exert control over the assessment centre when analysing the impact of environmental exposures on adverse health event outcomes in the UKB. Failure to control for this may lead to highly misleading results. Finally, the causal framework defined and data analyses support the role of personal characteristics and behavioural information as confounders in long-term air pollution analyses.

Chapter 14

Appendix

14.1 Cox model and the Nested case control (NCC): an empirical comparison

In this thesis I evaluated a potential and equivalently effective alternative to the Cox PH model to investigate long-term effect of air pollution on mortality. Briefly, I explored the use of the nested case-control design to investigate long-term effects of air pollution on health outcomes. Following a comparison analysis among the two methods, I opted for the use of the time-varying Cox PH model in the main analyses of this PhD project. The reasons that lead to this decision are shortly described in the rest of the section.

In environmental studies the event outcomes are traditionally investigated throughout the cohort designs. When performing longitudinal time-to-event studies, these practically consist of time-stratified comparisons between each case with the rest of the cohort of subjects was has not experienced yet the event under study. The Cox model is often applied when assessing relation-

ships between exposures and risks of events using hazard ratios (HRs) are reporting tools of the associations, along with the confidence bands. The cohort is therefore a type of case-control design in which the "risk set" of controls is represented by all the controls still in the follow-up. On the other hand, the nested case control is a specific case of the cohort design in which only a sample of controls is used to assess the association. The term "nested" is used to indicate that the sample is embedded within a larger prospective cohort and that the temporal relationship between exposure and outcome is thereby preserved.

In technical terms, the hazard function for the Cox proportional hazard (PH) model is known have the form:

$$h(y_i|\boldsymbol{\beta}) = h_0(y_i|\boldsymbol{\beta}) \exp(\boldsymbol{\beta}^T \mathbf{x}_i(t))$$

where

- $h_0(y_i|\beta)$: the unspecified baseline hazard function
- $i \in [1, n]$: each individual and n is the total number of individuals.
- $y_i = \min(t_i, c_i)$: t_i is time-to-event (failure time) and c_i is right-censoring time.
- $\mathbf{x}_i = (x_{i1}(t), x_{i2}(t), \ldots, x_{ip}(t))^T$: p-vector of potentially time-varying features for the individual i.
- $\boldsymbol{\beta} = (\beta_1, \beta_2, \ldots, \beta_p)^T$: p-vector of underlying model parameters.

The n observed data $\mathbf{D} = \{(y_i, \delta_i, \mathbf{x}_i) : i = 1, \ldots, n\}$, where $\delta_i = I(t_i \leq c_i)$ is an indicator variable such that $\delta_i = 1$ if the observation is not censored

and 0 otherwise.

14.1.1 Partial Likelihood

To estimate the underlying parameters $\boldsymbol{\beta}$, the original likelihood $L(\boldsymbol{\beta}|\mathbf{D})$ is hard to maximize. Cox proposed to maximize the partial likelihood:

$$L_p(\boldsymbol{\beta}|\mathbf{D}) = \prod_{i=1}^{n} \left(\frac{\exp(\boldsymbol{\beta}^T \mathbf{x}_i(t)}{\sum_{t \in R(y_i)} \exp(\boldsymbol{\beta}^T \mathbf{x}_i(t)} \right)^{\delta_i}$$

where $R(y_i)$ is the risk set of the i-th observation, defined as $R(y_i) = \{t : y_t \geq y_i\}$. In the cohort design $R(y_i)$ is represented by all the other participants still in study, while in the nested case control case $R(y_i)$ is a sample of them of size $|R(y_i)| = m$[1].

This equivalence states that, if the follow-up of a study is split in intervals defined by a time variable, then the Poisson model controlled for that variable and the Cox PH model in which the time variable is used as timescale, will produced identical results.

14.1.2 The case for the NCC

Imagine we are investigating the impact of air pollution exposure on all-cause mortality events using the UK Biobank. The cohort size is known to be approximately half a million subjects (n $\approx$ 500,000), with about 30,000 deaths occurring during a certain follow-up period and negligible loss to follow-up. This implies that, for each case and with a sufficiently fine time granularity (such as daily), the risk set $R(y_i)$ will be at least 470,000. Consequently, the entire analysis becomes unfeasible because the case-control dataset would

[1] If S is a finite set of subjects, $|S|$ is equal to the size of that set

contain at least 30,000 * 470,000 rows. Therefore, when dealing with large cohort databases of time-dependent exposures, it is common to aggregate the set of cases by specific time intervals. This is performed typically by calendar year and considering all the cases in a given year as tied events, which are then straightforwardly accounted for by tie methods implemented in every Cox PH routines (such as the Breslow or the Efron method).

An attractive alternative to using the entire $R(y_i)$ is to reduce the number of rows by selecting a sample of them, in other words to apply the NCC design. Theoretically, this method has been shown to be asimptotically consistent with the cohort design and therefore should lead to comparable, if not identical, estimates of association between exposure and outcomes[129]. For example, in our hypothetical scenario, reducing the number of controls to m=100 would result in 30,000 * 100 = 3 million rows, a sample size that is manageable by most modern computers. In environmental epidemiology, the NCC design was applied in a previous analysis of a Danish cohort which investigated the effect air pollutants on mortality by selecting 5 controls per case and assigning 5-year average prior to the year of the event.

14.1.3 Application on the UK Biobank database

The hypothetical scenario I described in the previous paragraph closely mirrors the actual investigation I conducted on the UK Biobank during my PhD. With resolved air pollution predictions available on a daily basis, I ideally should have been able to assign $PM_{2.5}$ averages on the same scale, such as by using 365-day averages backward from the exact day of the event. This approach would have represented the first attempt to assign exposure with such a high level of temporal resolution, compared to previous time-varying

studies. Unfortunately, as mentioned earlier, computational limitations even prevented the creation of the full survival dataset. Consequently, I opted for the application of the NCC design.

Because the NCC has been only applied once before in environmental epidemiology, before applying to my final analyses I endeavoured to compare its corresponding associations in the analysis between long-term $PM_{2.5}$ and non-accidental mortality with the estimates from a classical Cox PH model with follow-up split in yearly intervals. First, I randomly selected control subjects for each case, matching them within each risk set by calendar time. The literature does not specify the optimal number of controls to select, so I conducted an analysis using different sets of controls to assess the sensitivity of the results to the number of controls. I set m to 5, 10, and 100, and performed 50 re-sampling of the NCC dataset for each value of m. Within the Cox PH modelling framework, I applied a conditional logistic regression model to estimate the linear exposure-response association for each sample adjusting for age, assessment centre, sex, area deprivation, greenspace and urban-rural residential classification.

On the left side of fig 14.1 is it shown the distribution of the estimated HRs while on on right side it is shown the HR and 95% confidence interval from the yearly Cox PH model. In the NCC plot we can see that the distribution of the HR values increases with the number of controls, suggesting a correlation between the number of controls and the magnitude of the associations. This result is consistent with a methodological investigation which also found changes in the HRs when different number of controls were considered [130]. However, the NCC estimates on the left side were comparable to the Cox Model, showing almost identical hazard ratios (HR) to the NCC design with

100 controls.

Although not substantially different, this comparison led me to decide against using the NCC design due to its potential for biased health estimates that are dependent on the number of controls chosen, opting instead for the established Cox PH model with time-varying exposure and yearly aggregated data.

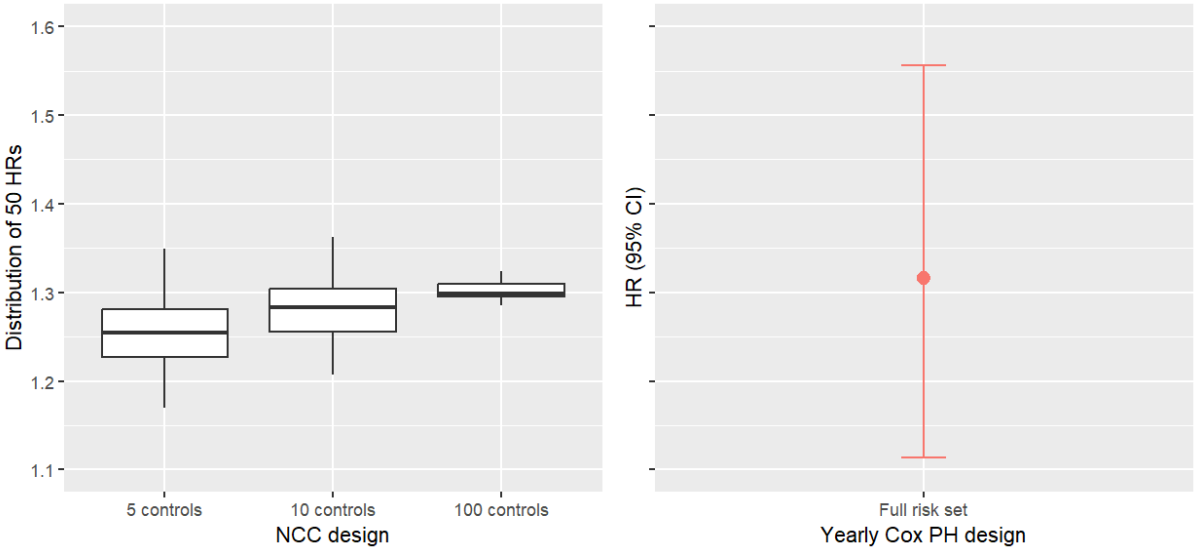


Figure 14.1: NCC - Cox PH comparison: the left-side panel shows the distribution of the HRs for non-accidental mortality from 50 sampled NCC datasets. On the right side panel, the HR (95% CI) from the yearly Cox PH model.

14.2 Spatio-Temporally reconstructed Daily $PM_{2.5}$ Concentrations across UK

The methodology used to model the novel exposure estimates is fully described in a previous publication [123] and it is independent of my project. The exposure model combines state-of-the-art ML techniques with a comprehensive environmental feature dataset to map daily levels of air pollutants

in Great Britain from 2003 to 2021. The target data are observations from ground PM monitors and the predictors include: AOD remote sensing satellite observations, modelled PM from UK chemical transport model, metereological and land variables (such as temperature, humidity, wind speed and elevation), vegetation index, population and road density, light data and distance measures from roadways, airports and seashore[29]. Each feature was harmonized to 1km resolution and extracted at monitoring sites to create the training dataset.

Models used single and ensemble-based algorithms featuring random forests (RF), extreme gradient boosting (XGB), light gradient boosting machine (LGBM), as well as lasso and ridge regression.

This model was applied in 4 stages. Briefly, Stage 1 augments monitoring station data using co-located monitors of PM10, while Stage 2 fills the gaps in satellites measurements using atmospheric re-analysis data. Stage 3 builds the ensemble spatio-temporal ML model of $PM_{2.5}$. Stage 4 applies the model to assign exposure values over the whole spatial and temporal domain.

Results show a good ensemble model performance, calculated through a ten-fold monitor-based cross-validation procedure, with an average R2 of 0.802 (0.746-0.888) for $PM_{2.5}$. Reconstructed pollution levels decreased markedly within the study period, with a stronger reduction in the latter eight years. The output generated from the ML ensemble models consists of over 5 billion data-points corresponding to 1kmx1km and daily resolved air pollution values, and it is unique in its combination of spatio-temporal resolution and coverage in the Great Britain. This data may contribute substantially to epidemiological research on the health effects of air pollution. The output predictions from the model are now linked to the UK Biobank cohort

database.

Bibliography

[1] World Health Organization. *WHO global air quality guidelines. Particulate matter (PM 2.5 and PM 10), ozone, nitrogen dioxide, sulfur dioxide and carbon monoxide. G.* Licence: CC BY-NC-SA 3.0 IG. 2021 (pages 12, 13).

[2] Lindsay Wichers Stanek and James S. Brown. "Air Pollution: Sources, Regulation, and Health Effects⋆". In: *Reference Module in Biomedical Sciences*. Elsevier, Jan. 2019. ISBN: 978-0-12-801238-3. DOI: 10.1016/B978-0-12-801238-3.11384-4. URL: https://www.sciencedirect.com/science/article/pii/B9780128012383113844 (visited on 07/15/2024) (page 12).

[3] *Air quality in Europe 2022*. Tech. rep. Report no. 05/2022. ISBN 978-92-9480-515-7. European Environment Agency (page 12).

[4] A. Jaworek, A. T. Sobczyk, A. Marchewicz, A. Krupa, and T. Czech. "Particulate matter emission control from small residential boilers after biomass combustion. A review". In: *Renewable and Sustainable Energy Reviews* 137 (Mar. 2021), p. 110446. ISSN: 1364-0321. DOI: 10.1016/j.rser.2020.110446. URL: https://www.sciencedirect.com/science/article/pii/S1364032120307334 (visited on 07/15/2024) (page 12).

[5] Jacopo Lunghi, Maurizio Malpede, and Lara Aleluia Reis. "Exploring the impact of livestock on air quality: A deep dive into Ammonia and particulate matter in Lombardy". In: *Environmental Impact Assessment Review* 105 (Mar. 2024), p. 107456. ISSN: 0195-9255. DOI: 10.1016/j.eiar.2024.107456. URL: https://www.sciencedirect.com/science/article/pii/S019592552400043X (visited on 07/15/2024) (page 12).

[6] Peter Faber, Frank Drewnick, and Stephan Borrmann. "Aerosol particle and trace gas emissions from earthworks, road construction, and asphalt paving in Germany: Emission factors and influence on local air quality". In: *Atmospheric Environment* 122 (Dec. 2015), pp. 662–671. ISSN: 1352-2310. DOI: 10.1016/j.atmosenv.2015.10.036. URL: https://www.sciencedirect.com/science/article/pii/S1352231015304544 (visited on 07/15/2024) (page 12).

[7] Sánchez Sorondo Wael K. Al-Delaimy Marcelo Sánchez Sorondo and Veerabhadran Ramanathan. *Health of People, Health of Planet and Our Responsibility*. Cham: Springer, 2020 (pages 13, 15).

[8] Muthiah Vaduganathan, George A. Mensah, Justine Varieur Turco, Valentin Fuster, and Gregory A. Roth. "The Global Burden of Cardiovascular Diseases and Risk". In: *Journal of the American College of Cardiology* 80.25 (Dec. 2022). Publisher: American College of Cardiology Foundation, pp. 2361–2371. DOI: 10.1016/j.jacc.2022.11.005. URL: https://www.jacc.org/doi/10.1016/j.jacc.2022.11.005 (visited on 01/15/2024) (page 13).

[9] Jelle Vlaanderen, Roel Vermeulen, Matthew Whitaker, Marc Chadeau-Hyam, Jouke-Jan Hottenga, Eco de Geus, Gonneke Willemsen, Brenda

W. J. H. Penninx, Rick Jansen, and Dorret I. Boomsma. "Impact of long-term exposure to PM2.5 on peripheral blood gene expression pathways involved in cell signaling and immune response". In: *Environment International* 168 (Oct. 2022), p. 107491. ISSN: 0160-4120. DOI: 10.1016/j.envint.2022.107491. URL: https://www.sciencedirect.com/science/article/pii/S0160412022004184 (visited on 06/19/2024) (page 13).

[10] Rajagopalan Sanjay, G. Al-Kindi Sadeer, and D. Brook Robert. "Air Pollution and Cardiovascular Disease: JACC State-of-the-Art Review". In: *Journal of the American College of Cardiology* 72.17 (2018), pp. 2054–2070. ISSN: 0735-1097. DOI: 10.1016/j.jacc.2018.07.099. URL: https://www.sciencedirect.com/science/article/pii/S0735109718383554 (page 13).

[11] Pascale S. J. Lakey, Thomas Berkemeier, Haijie Tong, Andrea M. Arangio, Kurt Lucas, Ulrich Pöschl, and Manabu Shiraiwa. "Chemical exposure-response relationship between air pollutants and reactive oxygen species in the human respiratory tract". en. In: *Scientific Reports* 6.1 (Sept. 2016). Publisher: Nature Publishing Group, p. 32916. ISSN: 2045-2322. DOI: 10.1038/srep32916. URL: https://www.nature.com/articles/srep32916 (visited on 07/15/2024) (page 13).

[12] Hermes Reyes-Caballero, Xiaoquan Rao, Qiushi Sun, Marc O. Warmoes, Penghui Lin, Tom E. Sussan, Bongsoo Park, Teresa W.-M. Fan, Andrei Maiseyeu, Sanjay Rajagopalan, Geoffrey D. Girnun, and Shyam Biswal. "Air pollution-derived particulate matter dysregulates hepatic Krebs cycle, glucose and lipid metabolism in mice". en. In: *Sci-*

entific Reports 9.1 (Nov. 2019). Publisher: Nature Publishing Group, p. 17423. ISSN: 2045-2322. DOI: 10.1038/s41598-019-53716-y. URL: https://www.nature.com/articles/s41598-019-53716-y (visited on 07/15/2024) (page 13).

[13] *Particulate Matter Induces Tissue OxInflammation: From Mechanism to Damage | Antioxidants & Redox Signaling*. URL: https://www.liebertpub.com/doi/10.1089/ars.2019.8015?url_ver=Z39.88-2003&rfr_id=ori%3Arid%3Acrossref.org&rfr_dat=cr_pub++0pubmed (visited on 07/15/2024) (page 14).

[14] *Air pollution and biomarkers of systemic inflammation and tissue repair in COPD patients | European Respiratory Society*. URL: https://erj.ersjournals.com/content/44/3/603 (visited on 06/19/2024) (page 14).

[15] Peter Libby, Paul M. Ridker, and Attilio Maseri. "Inflammation and Atherosclerosis". In: *Circulation* 105.9 (Mar. 2002). Publisher: American Heart Association, pp. 1135–1143. DOI: 10.1161/hc0902.104353. URL: https://www.ahajournals.org/doi/10.1161/hc0902.104353 (visited on 08/29/2024) (page 14).

[16] Christopher F. Rider and Chris Carlsten. "Air pollution and DNA methylation: effects of exposure in humans". In: *Clinical Epigenetics* 11.1 (Sept. 2019), p. 131. ISSN: 1868-7083. DOI: 10.1186/s13148-019-0713-2. URL: https://doi.org/10.1186/s13148-019-0713-2 (visited on 06/19/2024) (pages 14, 15).

[17] C. Arden Pope III. "What Do Epidemiologic Findings Tell Us about Health Effects of Environmental Aerosols?" In: *Journal of Aerosol Medicine* 13.4 (Jan. 2000). Publisher: Mary Ann Liebert, Inc., pub-

lishers, pp. 335–354. ISSN: 0894-2684. DOI: 10.1089/jam.2000.13.335. URL: https://www.liebertpub.com/doi/10.1089/jam.2000.13.335 (visited on 08/29/2024) (pages 14, 15).

[18] Frederick W. Lipfert. "Air Pollution and Mortality: Timing Is Everything". en. In: *Atmosphere* 11.12 (Dec. 2020). Number: 12 Publisher: Multidisciplinary Digital Publishing Institute, p. 1274. ISSN: 2073-4433. DOI: 10.3390/atmos11121274. URL: https://www.mdpi.com/2073-4433/11/12/1274 (visited on 08/29/2024) (pages 14, 32).

[19] Sadeer G. Al-Kindi, Robert D. Brook, Shyam Biswal, and Sanjay Rajagopalan. "Environmental determinants of cardiovascular disease: lessons learned from air pollution". en. In: *Nature Reviews Cardiology* 17.10 (Oct. 2020). Number: 10 Publisher: Nature Publishing Group, pp. 656–672. ISSN: 1759-5010. DOI: 10.1038/s41569-020-0371-2. URL: https://www.nature.com/articles/s41569-020-0371-2 (visited on 01/26/2024) (pages 14, 15).

[20] "Effect of air pollution on the human immune system". en. In: *Nature Medicine* 28.12 (Dec. 2022). Publisher: Nature Publishing Group, pp. 2482–2483. ISSN: 1546-170X. DOI: 10.1038/s41591-022-02093-7. URL: https://www.nature.com/articles/s41591-022-02093-7 (visited on 08/06/2024) (page 15).

[21] David E. Newby, Pier M. Mannucci, Grethe S. Tell, Andrea A. Baccarelli, Robert D. Brook, Ken Donaldson, Francesco Forastiere, Massimo Franchini, Oscar H. Franco, Ian Graham, Gerard Hoek, Barbara Hoffmann, Marc F. Hoylaerts, Nino Künzli, Nicholas Mills, Juha Pekkanen, Annette Peters, Massimo F. Piepoli, Sanjay Rajagopalan,

Robert F. Storey, and European Association for Cardiovascular Prevention and Rehabilitation and ESC Heart Failure Association on behalf of ESC Working Group on Thrombosis. "Expert position paper on air pollution and cardiovascular disease". In: *European Heart Journal* 36.2 (Jan. 2015), pp. 83–93. ISSN: 0195-668X. DOI: 10.1093/eurheartj/ehu458. URL: https://doi.org/10.1093/eurheartj/ehu458 (visited on 08/05/2024) (page 15).

[22] Loren E. Wold, Zhekang Ying, Kirk R. Hutchinson, Markus Velten, Matthew W. Gorr, Christina Velten, Dane J. Youtz, Aixia Wang, Pamela A. Lucchesi, Qinghua Sun, and Sanjay Rajagopalan. "Cardiovascular Remodeling in Response to Long-Term Exposure to Fine Particulate Matter Air Pollution". In: *Circulation: Heart Failure* 5.4 (July 2012). Publisher: American Heart Association, pp. 452–461. DOI: 10.1161/CIRCHEARTFAILURE.112.966580. URL: https://www.ahajournals.org/doi/10.1161/CIRCHEARTFAILURE.112.966580 (visited on 08/06/2024) (page 15).

[23] Daniel Krewski, Michael Jerrett, Richard Burnett, Renjun Ma, Edward Hughes, Yuanli Shi, Michelle Turner, C. Pope, George Thurston, Eugenia Calle, Michael Thun, Bernie Beckerman, Pat DeLuca, Norm Finkelstein, Kaz Ito, D. Moore, Bruce Newbold, Tim Ramsay, Zev Ross, and Barbara Tempalski. "Extended Follow-Up and Spatial Analysis of the American Cancer Society Study Linking Particulate Air Pollution and Mortality". In: *Research report (Health Effects Institute)* 140 (2009), 5–114, discussion 115 (pages 15, 17, 35).

[24] Douglas W. Dockery, C. Arden Pope, Xiping Xu, John D. Spengler, James H. Ware, Martha E. Fay, Benjamin G. Ferris, and Frank E.

Speizer. “An Association between Air Pollution and Mortality in Six U.S. Cities”. In: *New England Journal of Medicine* 329.24 (1993), pp. 1753–1759. DOI: 10.1056/NEJM199312093292401 (pages 15, 17, 20, 135).

[25] Douglas W. Dockery. “Health Effects of Particulate Air Pollution”. In: *Annals of Epidemiology* 19.4 (Apr. 2009), pp. 257–263. ISSN: 1047-2797. DOI: 10.1016/j.annepidem.2009.01.018. URL: https://www.sciencedirect.com/science/article/pii/S1047279709000337 (visited on 08/05/2024) (page 15).

[26] Stacey E. Alexeeff, Noelle S. Liao, Xi Liu, Stephen K. Van Den Eeden, and Stephen Sidney. “Long-Term PM2.5 Exposure and Risks of Ischemic Heart Disease and Stroke Events: Review and Meta-Analysis”. In: *Journal of the American Heart Association* 10.1 (Jan. 2021). Publisher: Wiley, e016890. DOI: 10.1161/JAHA.120.016890. URL: https://www.ahajournals.org/doi/10.1161/JAHA.120.016890 (visited on 02/01/2024) (pages 15, 16, 134).

[27] C. A. Pope 3rd, N. Coleman, Z. A. Pond, and R. T. Burnett. “Fine particulate air pollution and human mortality: 25+ years of cohort studies”. eng. In: *Environ Res* 183 (Apr. 2020). Edition: 20191114, p. 108924. ISSN: 0013-9351. DOI: 10.1016/j.envres.2019.108924 (pages 15, 19, 35, 132).

[28] Jie Chen and Gerard Hoek. “Long-term exposure to PM and all-cause and cause-specific mortality: A systematic review and meta-analysis”. In: *Environment International* 143 (2020), p. 105974. ISSN: 0160-4120. DOI: https://doi.org/10.1016/j.envint.2020.105974.

URL: https://www.sciencedirect.com/science/article/pii/S0160412020319292 (pages 15, 19, 33, 35, 132).

[29] Itai Kloog, Francesco Nordio, Brent Coull, and Joel Schwartz. "Incorporating Local Land Use Regression And Satellite Aerosol Optical Depth In A Hybrid Model Of Spatiotemporal PM2.5 Exposures In The Mid-Atlantic States". In: *Environmental science \& technology* 46 (2012). DOI: 10.1021/es302673e (pages 15, 20, 21, 150).

[30] Duan Rui-Rui, Hao Ke, and Yang Ting. "Air pollution and chronic obstructive pulmonary disease". In: *Chronic Diseases and Translational Medicine* 6.4 (2020), pp. 260–269. ISSN: 2095-882X. DOI: 10.1016/j.cdtm.2020.05.004. URL: https://www.sciencedirect.com/science/article/pii/S2095882X20300438 (page 15).

[31] Yutong Cai, Wilma Zijlema, Dany Doiron, Marta Blangiardo, Paul Burton, Isabel Fortier, Amadou Gaye, John Gulliver, Kees de Hoogh, Kristian Hveem, Stephane Mbatchou, David Morley, Paul Elliott, Anna Hansell, and Susan Hodgson. "Ambient air pollution, traffic noise and adult asthma prevalence: a BioSHaRE approach". In: *European Respiratory Journal* 49 (2016). DOI: 10.1183/13993003.02127-2015 (pages 15, 30).

[32] Yang Bo-Yi, Fan Shujun, Thiering Elisabeth, Seissler Jochen, Nowak Dennis, Dong Guang-Hui, and Heinrich Joachim. "Ambient air pollution and diabetes: A systematic review and meta-analysis". In: *Environmental Research* 180 (2020), p. 108817. ISSN: 0013-9351. DOI: 10.1016/j.envres.2019.108817. URL: https://www.sciencedirect.com/science/article/pii/S0013935119306140 (page 16).

[33] B. Hamra Ghassan, Guha Neela, Cohen Aaron, Laden Francine, Raaschou-Nielsen Ole, M. Samet Jonathan, Vineis Paolo, Forastiere Francesco, Saldiva Paulo, Yorifuji Takashi, and Loomis Dana. "Outdoor Particulate Matter Exposure and Lung Cancer: A Systematic Review and Meta-Analysis". In: *Environmental Health Perspectives* 122.9 (2014), pp. 906–911. DOI: 10.1289/ehp/1408092 (page 16).

[34] Pengfei Fu, Xinbiao Guo, Felix Man Ho Cheung, and Ken Kin Lam Yung. "The association between PM2.5 exposure and neurological disorders: A systematic review and meta-analysis". In: *Science of The Total Environment* 655 (2019), pp. 1240–1248. ISSN: 0048-9697. DOI: https://doi.org/10.1016/j.scitotenv.2018.11.218. URL: https://www.sciencedirect.com/science/article/pii/S0048969718345741 (page 16).

[35] Raymond Roy and Amedeo D'Angiulli. "Air pollution and neurological diseases, current state highlights". English. In: *Frontiers in Neuroscience* 18 (Mar. 2024). Publisher: Frontiers. ISSN: 1662-453X. DOI: 10.3389/fnins.2024.1351721. URL: https://www.frontiersin.org/journals/neuroscience/articles/10.3389/fnins.2024.1351721/full (visited on 06/19/2024) (page 16).

[36] D. Krewski, R. T. Burnett, M. Goldberg, K. Hoover, J. Siemiatycki, M. Abrahamowicz, and W. White. "Reanalysis of the Harvard Six Cities Study, Part I: Validation and Replication". In: *Inhalation Toxicology* 17.7-8 (Jan. 2005). Publisher: Taylor & Francis _eprint: https://doi.org/10.1080/08958370590929402, pp. 335–342. ISSN: 0895-8378. DOI: 10.1080/08958370590929402. URL: https://doi.

org/10.1080/08958370590929402 (visited on 08/29/2024) (pages 17, 135).

[37] C. A. rd Thun M. J. Namboodiri M. M. Dockery D. W. Evans J. S. Speizer F. E. Heath C. W. Pope Jr. "Particulate air pollution as a predictor of mortality in a prospective study of U.S. adults." In: *Am J Respir Crit Care Med.* 151 (1995), pp. 669–674. DOI: 10.1164/ajrccm/151.3_Pt_1.669 (pages 17, 20, 135).

[38] Beelen Rob, Raaschou-Nielsen Ole, Stafoggia Massimo, Andersen Zorana Jovanovic, Weinmayr Gudrun, Hoffmann Barbara, Wolf Kathrin, Samoli Evangelia, Fischer Paul, Nieuwenhuijsen Mark, Vineis Paolo, W. Xun Wei, Katsouyanni Klea, Dimakopoulou Konstantina, Oudin Anna, Forsberg Bertil, Modig Lars, S. Havulinna Aki, Lanki Timo, Turunen Anu, Oftedal Bente, Nystad Wenche, Nafstad Per, Ulf, L. Pedersen Nancy, Östenson Claes-Göran, Fratiglioni Laura, Penell Johanna, Korek Michal, Pershagen Göran, Eriksen Kirsten Thorup, Overvad Kim, Ellermann Thomas, Eeftens Marloes, H. Peeters Petra, Meliefste Kees, Wang Meng, Bueno-de-Mesquita Bas, Sugiri Dorothea, Krämer Ursula, Heinrich Joachim, Kees, Key Timothy, Peters Annette, Hampel Regina, Concin Hans, Nagel Gabriele, Ineichen Alex, Schaffner Emmanuel, Probst-Hensch Nicole, Künzli Nino, Schindler Christian, Schikowski Tamara, Adam Martin, Phuleria Harish, Vilier Alice, Clavel-Chapelon Françoise, Declercq Christophe, Grioni Sara, Krogh Vittorio, Tsai Ming-Yi, Ricceri Fulvio, Sacerdote Carlotta, Galassi Claudia, Migliore Enrica, Ranzi Andrea, Cesaroni Giulia, Badaloni Chiara, Forastiere Francesco, Tamayo Ibon, Amiano Pilar, Dorronsoro Miren, Katsoulis Michail, Trichopoulou Antonia, Brunekreef Bert, and Hoek

Gerard. "Effects of long-term exposure to air pollution on natural-cause mortality: an analysis of 22 European cohorts within the multicentre ESCAPE project". In: *The Lancet* 383.9919 (2014), pp. 785–795. ISSN: 0140-6736. DOI: 10.1016/S0140-6736(13)62158-3. URL: https://www.sciencedirect.com/science/article/pii/S0140673613621583 (pages 18, 34).

[39] Rob Beelen, Massimo Stafoggia, Ole Raaschou-Nielsen, Zorana Jovanovic Andersen, Wei W. Xun, Klea Katsouyanni, Konstantina Dimakopoulou, Bert Brunekreef, Gudrun Weinmayr, Barbara Hoffmann, Kathrin Wolf, Evangelia Samoli, Danny Houthuijs, Mark Nieuwenhuijsen, Anna Oudin, Bertil Forsberg, David Olsson, Veikko Salomaa, Timo Lanki, Tarja Yli-Tuomi, Bente Oftedal, Geir Aamodt, Per Nafstad, Ulf De Faire, Nancy L. Pedersen, Claes-Göran Östenson, Laura Fratiglioni, Johanna Penell, Michal Korek, Andrei Pyko, Kirsten Thorup Eriksen, Anne Tjønneland, Thomas Becker, Marloes Eeftens, Michiel Bots, Kees Meliefste, Meng Wang, Bas Bueno-de-Mesquita, Dorothea Sugiri, Ursula Krämer, Joachim Heinrich, Kees de Hoogh, Timothy Key, Annette Peters, Josef Cyrys, Hans Concin, Gabriele Nagel, Alex Ineichen, Emmanuel Schaffner, Nicole Probst-Hensch, Julia Dratva, Regina Ducret-Stich, Alice Vilier, Françoise Clavel-Chapelon, Morgane Stempfelet, Sara Grioni, Vittorio Krogh, Ming-Yi Tsai, Alessandro Marcon, Fulvio Ricceri, Carlotta Sacerdote, Claudia Galassi, Enrica Migliore, Andrea Ranzi, Giulia Cesaroni, Chiara Badaloni, Francesco Forastiere, Ibon Tamayo, Pilar Amiano, Miren Dorronsoro, Michail Katsoulis, Antonia Trichopoulou, Paolo Vineis, and Gerard Hoek. "Long-term Exposure to Air Pollution and Cardio-

vascular Mortality: An Analysis of 22 European Cohorts". en-US. In: *Epidemiology* 25.3 (May 2014), p. 368. ISSN: 1044-3983. DOI: 10.1097/EDE.0000000000000076. URL: https://journals.lww.com/epidem/Fulltext/2014/05000/Long_term_Exposure_to_Air_Pollution_and.8.aspx (visited on 06/01/2023) (pages 18, 135).

[40] Frederick W. Lipfert. "A critical review of the ESCAPE project for estimating long-term health effects of air pollution". en. In: *Environment International* 99 (Feb. 2017), pp. 87–96. ISSN: 0160-4120. DOI: 10.1016/j.envint.2016.11.028. URL: https://www.sciencedirect.com/science/article/pii/S0160412016309047 (visited on 11/10/2022) (pages 18, 35).

[41] Giulia Cesaroni, Chiara Badaloni, Claudio Gariazzo, Massimo Stafoggia, Roberto Sozzi, Marina Davoli, and Francesco Forastiere. "Long-Term Exposure to Urban Air Pollution and Mortality in a Cohort of More than a Million Adults in Rome". In: *Environmental Health Perspectives* 121.3 (Mar. 2013). Publisher: Environmental Health Perspectives, pp. 324–331. DOI: 10.1289/ehp.1205862. URL: https://ehp.niehs.nih.gov/doi/10.1289/ehp.1205862 (visited on 04/05/2023) (pages 18, 35).

[42] B. Hoffmann, G. Weinmayr, F. Hennig, K. Fuks, S. Moebus, C. Weimar, N. Dragano, D.M. Hermann, H. Kälsch, A.A. Mahabadi, R. Erbel, and K.H. Jöckel. "Air quality, stroke and coronary events-results of the Heinz Nixdorf Recall Study from the Ruhr region". In: *Deutsches Arzteblatt International* 112.12 (2015), pp. 195–201. DOI: 10.3238/arztebl.2015.0195 (page 18).

[43] Malek Bentayeb, Verene Wagner, Morgane Stempfelet, Marie Zins, Marcel Goldberg, Mathilde Pascal, Sophie Larrieu, Pascal Beaudeau, Sylvie Cassadou, Daniel Eilstein, Laurent Filleul, Alain Le Tertre, Sylvia Medina, Laurence Pascal, Helene Prouvost, Philippe Quénel, Abdelkrim Zeghnoun, and Agnes Lefranc. "Association between long-term exposure to air pollution and mortality in France: A 25-year follow-up study". In: *Environment International* 85 (Dec. 2015), pp. 5–14. ISSN: 0160-4120. DOI: 10.1016/j.envint.2015.08.006. URL: https://www.sciencedirect.com/science/article/pii/S0160412015300349 (visited on 07/17/2024) (page 18).

[44] Iain M. Carey, Richard W. Atkinson, Andrew J. Kent, Tjeerd van Staa, Derek G. Cook, and H. Ross Anderson. "Mortality Associations with Long-Term Exposure to Outdoor Air Pollution in a National English Cohort". In: *American Journal of Respiratory and Critical Care Medicine* 187.11 (June 2013). Publisher: American Thoracic Society - AJRCCM, pp. 1226–1233. ISSN: 1073-449X. DOI: 10.1164/rccm.201210-1758OC. URL: https://www.atsjournals.org/doi/10.1164/rccm.201210-1758OC (visited on 04/14/2023) (pages 18, 35, 133).

[45] B. Brunekreef and et al. *Mortality and Morbidity Effects of Long-Term Exposure to Low-Level PM2.5, BC, NO2, and O3: An Analysis of European Cohorts in the ELAPSE Project*. Tech. rep. Volume: Research Report 208. 2021 (pages 18, 20, 135).

[46] F. Dominici, J. Schwartz, Q. Di, D. Braun, C. Choirat, and A. Zanobetti. "Assessing Adverse Health Effects of Long-Term Exposure to Low Levels of Ambient Air Pollution: Phase 1". eng. In: *Res Rep Health Eff*

Inst 2019 (2019). Edition: 20191101. ISSN: 1041-5505 (Print) (pages 18, 25, 135).

[47] M. Brauer, J. R. Brook, T. Christidis, Y. Chu, D. L. Crouse, A. Erickson, P. Hystad, C. Li, R. V. Martin, J. Meng, A. J. Pappin, L. L. Pinault, M. Tjepkema, A. van Donkelaar, S. Weichenthal, and R. T. Burnett. "Mortality-Air Pollution Associations in Low-Exposure Environments (MAPLE): Phase 1". eng. In: *Res Rep Health Eff Inst* 203 (Nov. 2019), pp. 1–87. ISSN: 1041-5505 (Print) 1041-5505 (pages 18, 35, 135).

[48] Qian Di, Yan Wang, Antonella Zanobetti, Yun Wang, Petros Koutrakis, Christine Choirat, Francesca Dominici, and D. Schwartz Joel. "Air Pollution and Mortality in the Medicare Population". In: *New England Journal of Medicine* 376.26 (2017), pp. 2513–2522. DOI: 10.1056/NEJMoa1702747 (pages 18, 20, 22, 26, 135).

[49] Amanda J. Pappin, Tanya Christidis, Lauren L. Pinault, Dan L. Crouse, Jeffrey R. Brook, Anders Erickson, Perry Hystad, Chi Li, Randall V. Martin, Jun Meng, Scott Weichenthal, Aaron van Donkelaar, Michael Tjepkema, Michael Brauer, and Richard T. Burnett. "Examining the Shape of the Association between Low Levels of Fine Particulate Matter and Mortality across Three Cycles of the Canadian Census Health and Environment Cohort". In: *Environmental Health Perspectives* 127.10 (Oct. 2019). Publisher: Environmental Health Perspectives, p. 107008. DOI: 10.1289/EHP5204. URL: https://ehp.niehs.nih.gov/doi/full/10.1289/EHP5204 (visited on 07/24/2024) (pages 18, 20, 27, 33, 135).

[50] Tanya Christidis, Anders C. Erickson, Amanda J. Pappin, Daniel L. Crouse, Lauren L. Pinault, Scott A. Weichenthal, Jeffrey R. Brook, Aaron van Donkelaar, Perry Hystad, Randall V. Martin, Michael Tjepkema, Richard T. Burnett, and Michael Brauer. "Low concentrations of fine particle air pollution and mortality in the Canadian Community Health Survey cohort". In: *Environmental Health* 18.1 (Oct. 2019), p. 84. ISSN: 1476-069X. DOI: 10.1186/s12940-019-0518-y. URL: https://doi.org/10.1186/s12940-019-0518-y (visited on 06/28/2024) (pages 18, 33, 35).

[51] Maciej Strak, Gudrun Weinmayr, Sophia Rodopoulou, Jie Chen, Kees de Hoogh, Zorana J Andersen, Richard Atkinson, Mariska Bauwelinck, Terese Bekkevold, Tom Bellander, Marie-Christine Boutron-Ruault, Jørgen Brandt, Giulia Cesaroni, Hans Concin, Daniela Fecht, Francesco Forastiere, John Gulliver, Ole Hertel, Barbara Hoffmann, Ulla Arthur Hvidtfeldt, Nicole A H Janssen, Karl-Heinz Jöckel, Jeanette T Jørgensen, Matthias Ketzel, Jochem O Klompmaker, Anton Lager, Karin Leander, Shuo Liu, Petter Ljungman, Patrik K E Magnusson, Amar J Mehta, Gabriele Nagel, Bente Oftedal, Göran Pershagen, Annette Peters, Ole Raaschou-Nielsen, Matteo Renzi, Debora Rizzuto, Yvonne T van der Schouw, Sara Schramm, Gianluca Severi, Torben Sigsgaard, Mette Sørensen, Massimo Stafoggia, Anne Tjønneland, W M Monique Verschuren, Danielle Vienneau, Kathrin Wolf, Klea Katsouyanni, Bert Brunekreef, Gerard Hoek, and Evangelia Samoli. "Long term exposure to low level air pollution and mortality in eight European cohorts within the ELAPSE project: pooled analysis". In: *BMJ* 374 (Sept. 2021), n1904. DOI: 10.1136/bmj.n1904. URL: http://www.bmj.

com/content/374/bmj.n1904.abstract (pages 18–20, 24, 34, 35, 135).

[52] Ulla Arthur Hvidtfeldt, Jie Chen, Zorana Jovanovic Andersen, Richard Atkinson, Mariska Bauwelinck, Tom Bellander, Jørgen Brandt, Bert Brunekreef, Giulia Cesaroni, Hans Concin, Daniela Fecht, Francesco Forastiere, Carla H. van Gils, John Gulliver, Ole Hertel, Gerard Hoek, Barbara Hoffmann, Kees de Hoogh, Nicole Janssen, Jeanette Therming Jørgensen, Klea Katsouyanni, Karl-Heinz Jöckel, Matthias Ketzel, Jochem O. Klompmaker, Alois Lang, Karin Leander, Shuo Liu, Petter L. S. Ljungman, Patrik K. E. Magnusson, Amar Jayant Mehta, Gabriele Nagel, Bente Oftedal, Göran Pershagen, Raphael Simon Peter, Annette Peters, Matteo Renzi, Debora Rizzuto, Sophia Rodopoulou, Evangelia Samoli, Per Everhard Schwarze, Gianluca Severi, Torben Sigsgaard, Massimo Stafoggia, Maciej Strak, Danielle Vienneau, Gudrun Weinmayr, Kathrin Wolf, and Ole Raaschou-Nielsen. "Long-term exposure to fine particle elemental components and lung cancer incidence in the ELAPSE pooled cohort". eng. In: *Environmental research* 193 (Feb. 2021), p. 110568. ISSN: 0013-9351. DOI: 10.1016/j.envres.2020.110568. URL: http://europepmc.org/abstract/MED/33278469%20https://doi.org/10.1016/j.envres.2020.110568 (pages 18, 20).

[53] Shuo Liu, Jeanette T. Jørgensen, Petter Ljungman, Göran Pershagen, Tom Bellander, Karin Leander, Patrik K. E. Magnusson, Debora Rizzuto, Ulla A. Hvidtfeldt, Ole Raaschou-Nielsen, Kathrin Wolf, Barbara Hoffmann, Bert Brunekreef, Maciej Strak, Jie Chen, Amar Mehta, Richard W. Atkinson, Mariska Bauwelinck, Raphaëlle Varraso, Marie-

Christine Boutron-Ruault, Jørgen Brandt, Giulia Cesaroni, Francesco Forastiere, Daniela Fecht, John Gulliver, Ole Hertel, Kees de Hoogh, Nicole A. H. Janssen, Klea Katsouyanni, Matthias Ketzel, Jochem O. Klompmaker, Gabriele Nagel, Bente Oftedal, Annette Peters, Anne Tjønneland, Sophia P. Rodopoulou, Evangelia Samoli, Terese Bekkevold, Torben Sigsgaard, Massimo Stafoggia, Danielle Vienneau, Gudrun Weinmayr, Gerard Hoek, and Zorana J. Andersen. "Long-term exposure to low-level air pollution and incidence of chronic obstructive pulmonary disease: The ELAPSE project". In: *Environment International* 146 (Jan. 2021), p. 106267. ISSN: 0160-4120. DOI: 10.1016/j.envint.2020.106267. URL: https://www.sciencedirect.com/science/article/pii/S0160412020322224 (visited on 08/29/2024) (page 18).

[54] Ole Raaschou-Nielsen, Zorana J. Andersen, Rob Beelen, Evangelia Samoli, Massimo Stafoggia, Gudrun Weinmayr, Barbara Hoffmann, Paul Fischer, Mark J. Nieuwenhuijsen, Bert Brunekreef, Wei W. Xun, Klea Katsouyanni, Konstantina Dimakopoulou, Johan Sommar, Bertil Forsberg, Lars Modig, Anna Oudin, Bente Oftedal, Per E. Schwarze, Per Nafstad, Ulf De Faire, Nancy L. Pedersen, Claes-Göran Östenson, Laura Fratiglioni, Johanna Penell, Michal Korek, Göran Pershagen, Kirsten T. Eriksen, Mette Sørensen, Anne Tjønneland, Thomas Ellermann, Marloes Eeftens, Petra H. Peeters, Kees Meliefste, Meng Wang, Bas Bueno-de-Mesquita, Timothy J. Key, Kees de Hoogh, Hans Concin, Gabriele Nagel, Alice Vilier, Sara Grioni, Vittorio Krogh, Ming-Yi Tsai, Fulvio Ricceri, Carlotta Sacerdote, Claudia Galassi, Enrica Migliore, Andrea Ranzi, Giulia Cesaroni, Chiara Badaloni,

Francesco Forastiere, Ibon Tamayo, Pilar Amiano, Miren Dorronsoro, Antonia Trichopoulou, Christina Bamia, Paolo Vineis, and Gerard Hoek. "Air pollution and lung cancer incidence in 17 European cohorts: prospective analyses from the European Study of Cohorts for Air Pollution Effects (ESCAPE)". English. In: *The Lancet Oncology* 14.9 (Aug. 2013). Publisher: Elsevier, pp. 813–822. ISSN: 1470-2045, 1474-5488. DOI: 10.1016/S1470-2045(13)70279-1. URL: https://www.thelancet.com/journals/lanonc/article/PIIS1470-2045(13)70279-1/abstract (visited on 08/29/2024) (page 18).

[55] M. Stafoggia, B. Oftedal, J. Chen, S. Rodopoulou, M. Renzi, R. W. Atkinson, M. Bauwelinck, J. O. Klompmaker, A. Mehta, D. Vienneau, Z. J. Andersen, T. Bellander, J. Brandt, G. Cesaroni, K. de Hoogh, D. Fecht, J. Gulliver, O. Hertel, B. Hoffmann, U. A. Hvidtfeldt, K. H. Jöckel, J. T. Jørgensen, K. Katsouyanni, M. Ketzel, D. T. Kristoffersen, A. Lager, K. Leander, S. Liu, P. L. S. Ljungman, G. Nagel, G. Pershagen, A. Peters, O. Raaschou-Nielsen, D. Rizzuto, S. Schramm, P. E. Schwarze, G. Severi, T. Sigsgaard, M. Strak, Y. T. van der Schouw, M. Verschuren, G. Weinmayr, K. Wolf, E. Zitt, E. Samoli, F. Forastiere, B. Brunekreef, G. Hoek, and N. A. H. Janssen. "Long-term exposure to low ambient air pollution concentrations and mortality among 28 million people: results from seven large European cohorts within the ELAPSE project". eng. In: *Lancet Planet Health* 6.1 (Jan. 2022), e9–e18. ISSN: 2542-5196. DOI: 10.1016/s2542-5196(21)00277-1 (pages 18, 19, 24, 26, 32, 34, 142).

[56] Sophia Rodopoulou, Massimo Stafoggia, Jie Chen, Kees de Hoogh, Mariska Bauwelinck, Amar J. Mehta, Jochem O. Klompmaker, Bente

Oftedal, Danielle Vienneau, Nicole A. H. Janssen, Maciej Strak, Zorana J. Andersen, Matteo Renzi, Giulia Cesaroni, Carl Fredrik Nordheim, Terese Bekkevold, Richard Atkinson, Francesco Forastiere, Klea Katsouyanni, Bert Brunekreef, Evangelia Samoli, and Gerard Hoek. "Long-term exposure to fine particle elemental components and mortality in Europe: Results from six European administrative cohorts within the ELAPSE project". In: *Science of The Total Environment* 809 (Feb. 2022), p. 152205. ISSN: 0048-9697. DOI: 10.1016/j.scitotenv.2021.152205. URL: https://www.sciencedirect.com/science/article/pii/S0048969721072818 (visited on 08/29/2024) (pages 18, 34).

[57] Jie Chen, Danielle Braun, Tanya Christidis, Michael Cork, Sophia Rodopoulou, Evangelia Samoli, Massimo Stafoggia, Kathrin Wolf, Xiao Wu, Weiran Yuchi, Zorana J. Andersen, Richard Atkinson, Mariska Bauwelinck, Kees de Hoogh, Nicole A.H. Janssen, Klea Katsouyanni, Jochem O. Klompmaker, Doris Tove Kristoffersen, Youn-Hee Lim, Bente Oftedal, Maciej Strak, Danielle Vienneau, Jiawei Zhang, Richard T. Burnett, Gerard Hoek, Francesca Dominici, Michael Brauer, and Bert Brunekreef. "Long-Term Exposure to Low-Level PM2.5 and Mortality: Investigation of Heterogeneity by Harmonizing Analyses in Large Cohort Studies in Canada, United States, and Europe". In: *Environmental Health Perspectives* 131.12 (Dec. 2023). Publisher: Environmental Health Perspectives, p. 127003. DOI: 10.1289/EHP12141. URL: https://ehp.niehs.nih.gov/doi/full/10.1289/EHP12141 (visited on 02/22/2024) (pages 19, 32, 36).

[58] G. Hoek. "Methods for Assessing Long-Term Exposures to Outdoor Air Pollutants". eng. In: *Curr Environ Health Rep* 4.4 (Dec. 2017), pp. 450–462. ISSN: 2196-5412. DOI: 10.1007/s40572-017-0169-5 (pages 20, 21).

[59] Qian Di, Itai Kloog, Petros Koutrakis, Alexei Lyapustin, Yujie Wang, and Joel Schwartz. "Assessing PM2.5 Exposures with High Spatiotemporal Resolution across the Continental United States". In: *Environmental Science \& Technology* 50.9 (2016), pp. 4712–4721. DOI: 10.1021/acs.est.5b06121 (pages 20, 21, 32, 33).

[60] Marloes Eeftens, Rob Beelen, Kees de Hoogh, Tom Bellander, Giulia Cesaroni, Marta Cirach, Christophe Declercq, Audrius Dėdelė, Evi Dons, Audrey de Nazelle, Konstantina Dimakopoulou, Kirsten Eriksen, Grégoire Falq, Paul Fischer, Claudia Galassi, Regina Gražulevičienė, Joachim Heinrich, Barbara Hoffmann, Michael Jerrett, Dirk Keidel, Michal Korek, Timo Lanki, Sarah Lindley, Christian Madsen, Anna Mölter, Gizella Nádor, Mark Nieuwenhuijsen, Michael Nonnemacher, Xanthi Pedeli, Ole Raaschou-Nielsen, Evridiki Patelarou, Ulrich Quass, Andrea Ranzi, Christian Schindler, Morgane Stempfelet, Euripides Stephanou, Dorothea Sugiri, Ming-Yi Tsai, Tarja Yli-Tuomi, Mihály J. Varró, Danielle Vienneau, Stephanie von Klot, Kathrin Wolf, Bert Brunekreef, and Gerard Hoek. "Development of Land Use Regression Models for PM2.5, PM2.5 Absorbance, PM10 and PMcoarse in 20 European Study Areas; Results of the ESCAPE Project". In: *Environmental Science \& Technology* 46.20 (2012), pp. 11195–11205. DOI: 10.1021/es301948k (pages 20, 30, 32).

[61] Rob Beelen, Gerard Hoek, Danielle Vienneau, Marloes Eeftens, Konstantina Dimakopoulou, Xanthi Pedeli, Ming-Yi Tsai, Nino Künzli, Tamara Schikowski, Alessandro Marcon, Kirsten T. Eriksen, Ole Raaschou-Nielsen, Euripides Stephanou, Evridiki Patelarou, Timo Lanki, Tarja Yli-Tuomi, Christophe Declercq, Grégoire Falq, Morgane Stempfelet, Matthias Birk, Josef Cyrys, Stephanie von Klot, Gizella Nádor, Mihály János Varró, Audrius Dėdelė, Regina Gražulevičienė, Anna Möller, Sarah Lindley, Christian Madsen, Giulia Cesaroni, Andrea Ranzi, Chiara Badaloni, Barbara Hoffmann, Michael Nonnemacher, Ursula Krämer, Thomas Kuhlbusch, Marta Cirach, Audrey de Nazelle, Mark Nieuwenhuijsen, Tom Bellander, Michal Korek, David Olsson, Magnus Strömgren, Evi Dons, Michael Jerrett, Paul Fischer, Meng Wang, Bert Brunekreef, and Kees de Hoogh. "Development of NO2 and NOx land use regression models for estimating air pollution exposure in 36 study areas in Europe – The ESCAPE project". In: *Atmospheric Environment* 72 (June 2013), pp. 10–23. ISSN: 1352-2310. DOI: 10.1016/j.atmosenv.2013.02.037. URL: https://www.sciencedirect.com/science/article/pii/S1352231013001386 (pages 20, 30).

[62] Xuying Ma, Bin Zou, Jun Deng, Jay Gao, Ian Longley, Shun Xiao, Bin Guo, Yarui Wu, Tingting Xu, Xin Xu, Xiaosha Yang, Xiaoqi Wang, Zelei Tan, Yifan Wang, Lidia Morawska, and Jennifer Salmond. "A comprehensive review of the development of land use regression approaches for modeling spatiotemporal variations of ambient air pollution: A perspective from 2011 to 2023". In: *Environment International* 183 (Jan. 2024), p. 108430. ISSN: 0160-4120. DOI: 10.1016/j.envint.2024.108430. URL: https://www.sciencedirect.

com / science / article / pii / S0160412024000163 (visited on 08/29/2024) (page 20).

[63] N. S. Holmes and L. Morawska. "A review of dispersion modelling and its application to the dispersion of particles: An overview of different dispersion models available". In: *Atmospheric Environment* 40.30 (Sept. 2006), pp. 5902–5928. ISSN: 1352-2310. DOI: 10.1016/j.atmosenv.2006.06.003. URL: https://www.sciencedirect.com/science/article/pii/S1352231006006339 (visited on 08/29/2024) (page 20).

[64] Michael Jerrett, Altaf Arain, Pavlos Kanaroglou, Bernardo Beckerman, Dimitri Potoglou, Talar Sahsuvaroglu, Jason Morrison, and Chris Giovis. "A review and evaluation of intraurban air pollution exposure models". en. In: *Journal of Exposure Science & Environmental Epidemiology* 15.2 (Mar. 2005). Publisher: Nature Publishing Group, pp. 185–204. ISSN: 1559-064X. DOI: 10.1038/sj.jea.7500388. URL: https://www.nature.com/articles/7500388 (visited on 06/20/2024) (page 21).

[65] Q. Di, H. Amini, L. Shi, I. Kloog, R. Silvern, J. Kelly, M. B. Sabath, C. Choirat, P. Koutrakis, A. Lyapustin, Y. Wang, L. J. Mickley, and J. Schwartz. "An ensemble-based model of PM(2.5) concentration across the contiguous United States with high spatiotemporal resolution". eng. In: *Environ Int* 130 (Sept. 2019). Edition: 20190701, p. 104909. ISSN: 0160-4120 (Print) 0160-4120. DOI: 10.1016/j.envint.2019.104909 (page 21).

[66] J. Meng, C. Li, R. V. Martin, A. van Donkelaar, P. Hystad, and M. Brauer. "Estimated Long-Term (1981-2016) Concentrations of Am-

bient Fine Particulate Matter across North America from Chemical Transport Modeling, Satellite Remote Sensing, and Ground-Based Measurements". eng. In: *Environ Sci Technol* 53.9 (May 2019). Edition: 20190424, pp. 5071–5079. ISSN: 0013-936x. DOI: 10.1021/acs.est.8b06875 (pages 21, 32).

[67] Rochelle Schneider, Ana M. Vicedo-Cabrera, Francesco Sera, Pierre Masselot, Massimo Stafoggia, Kees de Hoogh, Itai Kloog, Stefan Reis, Massimo Vieno, and Antonio Gasparrini. "A Satellite-Based Spatio-Temporal Machine Learning Model to Reconstruct Daily PM2.5 Concentrations across Great Britain". In: *Remote Sensing* 12.22 (2020), p. 3803. ISSN: 2072-4292. URL: https://www.mdpi.com/2072-4292/12/22/3803 (page 21).

[68] Alonso Bussalleu, Gerard Hoek, Itai Kloog, Nicole Probst-Hensch, Martin Röösli, and Kees de Hoogh. "Modelling Europe-wide fine resolution daily ambient temperature for 2003–2020 using machine learning". In: *Science of The Total Environment* 928 (June 2024), p. 172454. ISSN: 0048-9697. DOI: 10.1016/j.scitotenv.2024.172454. URL: https://www.sciencedirect.com/science/article/pii/S0048969724026007 (visited on 07/22/2024) (page 21).

[69] Siddhartha Mandal, Ajit Rajiva, Itai Kloog, Jyothi S Menon, Kevin J Lane, Heresh Amini, Gagandeep K Walia, Shweta Dixit, Amruta Nori-Sarma, Anubrati Dutta, Praggya Sharma, Suganthi Jaganathan, Kishore K Madhipatla, Gregory A Wellenius, Jeroen de Bont, Chandra Venkataraman, Dorairaj Prabhakaran, Poornima Prabhakaran, Petter Ljungman, and Joel Schwartz. "Nationwide estimation of daily ambient PM2.5 from 2008 to 2020 at 1 km2 in India using an en-

semble approach". In: *PNAS Nexus* 3.3 (Mar. 2024), pgae088. ISSN: 2752-6542. DOI: 10.1093/pnasnexus/pgae088. URL: https://doi.org/10.1093/pnasnexus/pgae088 (visited on 07/22/2024) (page 21).

[70] John R. Nuckols, Mary H. Ward, and Lars Jarup. "Using Geographic Information Systems for Exposure Assessment in Environmental Epidemiology Studies". In: *Environmental Health Perspectives* 112.9 (2004), pp. 1007–1015. DOI: 10.1289/ehp.6738. URL: https://ehp.niehs.nih.gov/doi/abs/10.1289/ehp.6738%20%X%20Geographic%20information%20systems%20(GIS)%20are%20being%20used%20with%20increasing%20frequency%20in%20environmental%20epidemiology%20studies.%20Reported%20applications%20include%20locating%20the%20study%20population%20by%20geocoding%20addresses%20(assigning%20mapping%20coordinates),%20using%20proximity%20analysis%20of%20contaminant%20source%20as%20a%20surrogate%20for%20exposure,%20and%20integrating%20environmental%20monitoring%20data%20into%20the%20analysis%20of%20the%20health%20outcomes.%20Although%20most%20of%20these%20studies%20have%20been%20ecologic%20in%20design,%20some%20have%20used%20GIS%20in%20estimating%20environmental%20levels%20of%20a%20contaminant%20at%20the%20individual%20level%20and%20to%20design%20exposure%20metrics%20for%20use%20in%20epidemiologic%20studies.%20In%20this%20article%20we%20discuss%20fundamentals%20of%20three%20scientific%20disciplines%20instrumental%20to%20using%20GIS%20in%

20exposure%20assessment%20for%20epidemiologic%20studies:%20geospatial%20science,%20environmental%20science,%20and%20epidemiology.%20We%20also%20explore%20how%20a%20GIS%20can%20be%20used%20to%20accomplish%20several%20steps%20in%20the%20exposure%20assessment%20process.%20These%20steps%20include%20defining%20the%20study%20population,%20identifying%20source%20and%20potential%20routes%20of%20exposure,%20estimating%20environmental%20levels%20of%20target%20contaminants,%20and%20estimating%20personal%20exposures.%20We%20present%20and%20discuss%20examples%20for%20the%20first%20three%20steps.%20We%20discuss%20potential%20use%20of%20GIS%20and%20global%20positioning%20systems%20(GPS)%20in%20the%20last%20step.%20On%20the%20basis%20of%20our%20findings,%20we%20conclude%20that%20the%20use%20of%20GIS%20in%20exposure%20assessment%20for%20environmental%20epidemiology%20studies%20is%20not%20only%20feasible%20but%20can%20enhance%20the%20understanding%20of%20the%20association%20between%20contaminants%20in%20our%20environment%20and%20disease. (page 22).

[71] Terry M. Therneau and Patricia M. Grambsch. *Modeling Survival Data: Extending the Cox Model*. Ed. by K. Dietz, M. Gail, K. Krickeberg, J. Samet, and A. Tsiatis. Statistics for Biology and Health. New York, NY: Springer, 2000. ISBN: 978-1-4419-3161-0 978-1-4757-3294-8. DOI: 10.1007/978-1-4757-3294-8. URL: http://link.

springer.com/10.1007/978-1-4757-3294-8 (visited on 03/18/2024) (page 23).

[72] Francesca Dominici, Lianne Sheppard, and Merlise Clyde. "Health Effects of Air Pollution: A Statistical Review". In: *International Statistical Review* 71.2 (2003), pp. 243–276. DOI: 10.1111/j.1751-5823.2003.tb00195.x. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1751-5823.2003.tb00195.x (page 23).

[73] Douglas W. Dockery. "Synergy of Biostatistics and Epidemiology in Air Pollution Health Effects Studies". en. In: *International Statistical Review* 90.S1 (2022). _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/in S67–S81. ISSN: 1751-5823. DOI: 10.1111/insr.12525. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/insr.12525 (visited on 07/16/2024) (page 23).

[74] P. K. Andersen and R. D. Gill. "Cox's Regression Model for Counting Processes: A Large Sample Study". In: *The Annals of Statistics* 10.4 (1982), pp. 1100–1120, 21. URL: https://doi.org/10.1214/aos/1176345976 (pages 23, 32).

[75] Antonio Gasparrini. "Modeling exposure–lag–response associations with distributed lag non-linear models". In: *Statistics in Medicine* 33.5 (2014), pp. 881–899. DOI: 10.1002/sim.5963 (pages 24, 25).

[76] Hedi Katre Kriit, Eva M. Andersson, Hanne K. Carlsen, Niklas Andersson, Petter L. S. Ljungman, Göran Pershagen, David Segersson, Kristina Eneroth, Lars Gidhagen, Mårten Spanne, Peter Molnar, Patrik Wennberg, Annika Rosengren, Debora Rizzuto, Karin Leander, Diego Yacamán-Méndez, Patrik K. E. Magnusson, Bertil Forsberg, Leo Stockfelt, and Johan N. Sommar. "Using Distributed Lag Non-

Linear Models to Estimate Exposure Lag-Response Associations between Long-Term Air Pollution Exposure and Incidence of Cardiovascular Disease". en. In: *International Journal of Environmental Research and Public Health* 19.5 (Jan. 2022). Number: 5 Publisher: Multidisciplinary Digital Publishing Institute, p. 2630. ISSN: 1660-4601. DOI: 10.3390/ijerph19052630. URL: https://www.mdpi.com/1660-4601/19/5/2630 (visited on 06/01/2023) (pages 24, 25).

[77] Yaguang Wei, Yijing Feng, Mahdieh Danesh Yazdi, Kanhua Yin, Edgar Castro, Alexandra Shtein, Xinye Qiu, Adjani A. Peralta, Brent A. Coull, Francesca Dominici, and Joel D. Schwartz. "Exposure-response associations between chronic exposure to fine particulate matter and risks of hospital admission for major cardiovascular diseases: population based cohort study". en. In: *BMJ* 384 (Feb. 2024). Publisher: British Medical Journal Publishing Group Section: Research, e076939. ISSN: 1756-1833. DOI: 10.1136/bmj-2023-076939. URL: https://www.bmj.com/content/384/bmj-2023-076939 (visited on 07/26/2024) (page 25).

[78] Mahdieh Danesh Yazdi, Yan Wang, Qian Di, Weeberb J. Requia, Yaguang Wei, Liuhua Shi, Matthew Benjamin Sabath, Francesca Dominici, Brent Coull, John S. Evans, Petros Koutrakis, and Joel D. Schwartz. "Long-term effect of exposure to lower concentrations of air pollution on mortality among US Medicare participants and vulnerable subgroups: a doubly-robust approach". English. In: *The Lancet Planetary Health* 5.10 (Oct. 2021). Publisher: Elsevier, e689–e697. ISSN: 2542-5196. DOI: 10.1016/S2542-5196(21)00204-7. URL: https://www.thelancet.com/journals/lanplh/article/

PIIS2542-5196(21)00204-7/fulltext (visited on 05/29/2023) (pages 25, 141).

[79] Tyler J. VanderWeele and Ilya Shpitser. "On the definition of a confounder". In: *The Annals of Statistics* 41.1 (Feb. 2013). Publisher: Institute of Mathematical Statistics, pp. 196–220. ISSN: 0090-5364, 2168-8966. DOI: 10.1214/12-AOS1058. URL: https://projecteuclid.org/journals/annals-of-statistics/volume-41/issue-1/On-the-definition-of-a-confounder/10.1214/12-AOS1058.full (visited on 08/29/2024) (page 25).

[80] Lianne Sheppard, Richard T. Burnett, Adam A. Szpiro, Sun-Young Kim, Michael Jerrett, C. Arden Pope, and Bert Brunekreef. "Confounding and exposure measurement error in air pollution epidemiology". en. In: *Air Quality, Atmosphere & Health* 5.2 (June 2012), pp. 203–216. ISSN: 1873-9326. DOI: 10.1007/s11869-011-0140-9. URL: https://doi.org/10.1007/s11869-011-0140-9 (visited on 01/15/2024) (pages 25, 26, 35).

[81] Tyler J. VanderWeele. "Principles of confounder selection". en. In: *European Journal of Epidemiology* 34.3 (Mar. 2019), pp. 211–219. ISSN: 1573-7284. DOI: 10.1007/s10654-019-00494-6. URL: https://doi.org/10.1007/s10654-019-00494-6 (visited on 08/29/2024) (pages 25, 27).

[82] Anjum Hajat, Richard F. MacLehose, Anna Rosofsky, Katherine D. Walker, and Jane E. Clougherty. "Confounding by Socioeconomic Status in Epidemiological Studies of Air Pollution and Health: Challenges and Opportunities". In: *Environmental Health Perspectives* 129.6 (June 2021). Publisher: Environmental Health Perspectives, p. 065001.

DOI: 10.1289/EHP7980. URL: https://ehp.niehs.nih.gov/doi/10.1289/EHP7980 (visited on 05/23/2024) (pages 26, 27).

[83] Ki-Do Eum, Helen H. Suh, Vivian Chit Pun, and Justin Manjourides. "Impact of long-term temporal trends in fine particulate matter (PM2.5) on associations of annual PM2.5 exposure and mortality: An analysis of over 20 million Medicare beneficiaries". en-US. In: *Environmental Epidemiology* 2.2 (June 2018), e009. DOI: 10.1097/EE9.0000000000000009. URL: https://journals.lww.com/environepidem/fulltext/2018/06000/impact_of_long_term_temporal_trends_in_fine.1.aspx (visited on 08/29/2024) (pages 26, 33).

[84] Chi Li, Aaron van Donkelaar, Melanie S. Hammer, Erin E. McDuffie, Richard T. Burnett, Joseph V. Spadaro, Deepangsu Chatterjee, Aaron J. Cohen, Joshua S. Apte, Veronica A. Southerland, Susan C. Anenberg, Michael Brauer, and Randall V. Martin. "Reversal of trends in global fine particulate matter air pollution". en. In: *Nature Communications* 14.1 (Sept. 2023). Number: 1 Publisher: Nature Publishing Group, p. 5349. ISSN: 2041-1723. DOI: 10.1038/s41467-023-41086-z. URL: https://www.nature.com/articles/s41467-023-41086-z (visited on 01/15/2024) (pages 26, 32).

[85] Marc G. Weisskopf and Thomas F. Webster. "Trade-offs of Personal Versus More Proxy Exposure Measures in Environmental Epidemiology". en-US. In: *Epidemiology* 28.5 (Sept. 2017), p. 635. ISSN: 1044-3983. DOI: 10.1097/EDE.0000000000000686. URL: https://journals.lww.com/epidem/Fulltext/2017/09000/Trade_offs_of_Personal_Versus_More_Proxy_Exposure.2.aspx (visited on 05/29/2023) (page 26).

[86] Marie-Abèle Bind. "Causal Modeling in Environmental Health". In: *Annual Review of Public Health* 40.1 (2019), pp. 23–43. DOI: 10.1146/annurev-publhealth-040218-044048 (pages 27, 35, 141).

[87] Bernardo S. Beckerman, Michael Jerrett, Murray Finkelstein, Pavlos Kanaroglou, Jeffrey R. Brook, M. Altaf Arain, Malcolm R. Sears, David Stieb, John Balmes, and Kenneth Chapman. "The Association Between Chronic Exposure to Traffic-Related Air Pollution and Ischemic Heart Disease". In: *Journal of Toxicology and Environmental Health, Part A* 75.7 (Apr. 2012). Publisher: Taylor & Francis _eprint: https://doi.org/10.1080/15287394.2012.670899, pp. 402–411. ISSN: 1528-7394. DOI: 10.1080/15287394.2012.670899. URL: https://doi.org/10.1080/15287394.2012.670899 (visited on 07/24/2024) (page 27).

[88] W. Dana Flanders, Mitchel Klein, Lyndsey A. Darrow, Matthew J. Strickland, Stefanie E. Sarnat, Jeremy A. Sarnat, Lance A. Waller, Andrea Winquist, and Paige E. Tolbert. "A Method for Detection of Residual Confounding in Time-series and Other Observational Studies". en-US. In: *Epidemiology* 22.1 (Jan. 2011), p. 59. ISSN: 1044-3983. DOI: 10.1097/EDE.0b013e3181fdcabe. URL: https://journals.lww.com/epidem/fulltext/2011/01000/a_method_for_detection_of_residual_confounding_in.10.aspx (visited on 07/24/2024) (page 27).

[89] Marc G. Weisskopf, Marianthi-Anna Kioumourtzoglou, and Andrea L. Roberts. "Air Pollution and Autism Spectrum Disorders: Causal or Confounded?" en. In: *Current Environmental Health Reports* 2.4 (Dec. 2015), pp. 430–439. ISSN: 2196-5412. DOI: 10.1007/s40572-

015-0073-9. URL: https://doi.org/10.1007/s40572-015-0073-9 (visited on 07/24/2024) (pages 27, 34).

[90] *Use of directed acyclic graphs (DAGs) to identify confounders in applied health research: review and recommendations | International Journal of Epidemiology | Oxford Academic*. URL: https://academic.oup.com/ije/article/50/2/620/6012812 (visited on 07/26/2024) (page 27).

[91] C. Sudlow, J. Gallacher, N. Allen, V. Beral, P. Burton, J. Danesh, P. Downey, P. Elliott, J. Green, M. Landray, B. Liu, P. Matthews, G. Ong, J. Pell, A. Silman, A. Young, T. Sprosen, T. Peakman, and R. Collins. "UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age". eng. In: *PLoS Med* 12.3 (Mar. 2015). Edition: 2015/04/01, e1001779. ISSN: 1549-1277 (Print) 1549-1277. DOI: 10.1371/journal.pmed.1001779 (pages 28, 141).

[92] A. Fry, T. J. Littlejohns, C. Sudlow, N. Doherty, L. Adamska, T. Sprosen, R. Collins, and N. E. Allen. "Comparison of Sociodemographic and Health-Related Characteristics of UK Biobank Participants With Those of the General Population". eng. In: *Am J Epidemiol* 186.9 (Nov. 2017). Edition: 2017/06/24, pp. 1026–1034. ISSN: 1476-6256 (Electronic) 0002-9262 (Linking). DOI: 10.1093/aje/kwx246. URL: https://www.ncbi.nlm.nih.gov/pubmed/28641372 (page 28).

[93] Dany Doiron, Kees de Hoogh, Nicole Probst-Hensch, Isabel Fortier, Yutong Cai, Sara De Matteis, and Anna L. Hansell. "Air pollution, lung function and COPD: results from the population-based UK Biobank

study". en. In: *European Respiratory Journal* 54.1 (July 2019). Publisher: European Respiratory Society Section: Original Articles. ISSN: 0903-1936, 1399-3003. DOI: 10.1183/13993003.02140-2018. URL: https://erj.ersjournals.com/content/54/1/1802140 (visited on 06/21/2024) (page 30).

[94] Yutong Cai, Susan Hodgson, Marta Blangiardo, John Gulliver, David Morley, Daniela Fecht, Danielle Vienneau, Kees de Hoogh, Tim Key, Kristian Hveem, Paul Elliott, and Anna L. Hansell. "Road traffic noise, air pollution and incident cardiovascular disease: A joint analysis of the HUNT, EPIC-Oxford and UK Biobank cohorts". In: *Environment International* 114 (May 2018), pp. 191–201. ISSN: 0160-4120. DOI: 10.1016/j.envint.2018.02.048. URL: https://www.sciencedirect.com/science/article/pii/S0160412017320548 (visited on 02/01/2024) (page 30).

[95] Aung Nay, M. Sanghvi Mihir, Zemrak Filip, M. Lee Aaron, A. Cooper Jackie, M. Paiva Jose, J. Thomson Ross, Fung Kenneth, Y. Khanji Mohammed, Lukaschuk Elena, Carapella Valentina, Kim Young Jin, B. Munroe Patricia, K. Piechnik Stefan, Neubauer Stefan, and E. Petersen Steffen. "Association Between Ambient Air Pollution and Cardiac Morpho-Functional Phenotypes". In: *Circulation* 138.20 (2018), pp. 2175–2186. DOI: 10.1161/CIRCULATIONAHA.118.034856 (page 30).

[96] Breda Cullen, Danielle Newby, Duncan Lee, Donald Lyall, Alejo Nevado-Holgado, Jonathan Evans, Simon Lovestone, and Jonathan Cavanagh. "Cross-sectional and longitudinal analyses of outdoor air pollution exposure and cognitive function in UK Biobank". In: *Scientific Reports* 8 (2018). DOI: 10.1038/s41598-018-30568-6 (pages 30, 140).

[97] Dawson Hedges, Lance Erickson, Jackie Kunzelman, Bruce Brown, and Shawn Gale. "Association between Exposure to Air Pollution and Hippocampal Volume in Adults in the UK Biobank". In: *NeuroToxicology* 74 (2019). DOI: 10.1016/j.neuro.2019.06.005 (page 30).

[98] Sharon Y. L. Chua, Anthony P. Khawaja, James Morgan, Nicholas Strouthidis, Charles Reisman, Andrew D. Dick, Peng T. Khaw, Praveen J. Patel, Paul J. Foster, and for the UK Biobank Eye and Vision Consortium. "The Relationship Between Ambient Atmospheric Fine Particulate Matter (PM2.5) and Glaucoma in a Large Community Cohort". In: *Investigative Ophthalmology & Visual Science* 60.14 (Nov. 2019), pp. 4915–4923. ISSN: 1552-5783. DOI: 10.1167/iovs.19-28346. URL: https://doi.org/10.1167/iovs.19-28346 (visited on 06/21/2024) (page 30).

[99] Sharon Y. L. Chua, Anthony P. Khawaja, Andrew D. Dick, James Morgan, Baljean Dhillon, Andrew J. Lotery, Nicholas G. Strouthidis, Charles Reisman, Tunde Peto, Peng T. Khaw, Paul J. Foster, Praveen J. Patel, and on behalf of The UK Biobank Eye and Vision Consortium. "Ambient Air Pollution Associations with Retinal Morphology in the UK Biobank". In: *Investigative Ophthalmology & Visual Science* 61.5 (May 2020), p. 32. ISSN: 1552-5783. DOI: 10.1167/iovs.61.5.32. URL: https://doi.org/10.1167/iovs.61.5.32 (visited on 06/21/2024) (page 30).

[100] Jie Chen, Kees de Hoogh, John Gulliver, Barbara Hoffmann, Ole Hertel, Matthias Ketzel, Gudrun Weinmayr, Mariska Bauwelinck, Aaron van Donkelaar, Ulla A. Hvidtfeldt, Richard Atkinson, Nicole A. H. Janssen, Randall V. Martin, Evangelia Samoli, Zorana J. Andersen,

Bente M. Oftedal, Massimo Stafoggia, Tom Bellander, Maciej Strak, Kathrin Wolf, Danielle Vienneau, Bert Brunekreef, and Gerard Hoek. "Development of Europe-Wide Models for Particle Elemental Composition Using Supervised Linear Regression and Random Forest". In: *Environmental Science & Technology* 54.24 (Dec. 2020). Publisher: American Chemical Society, pp. 15698–15709. ISSN: 0013-936X. DOI: 10.1021/acs.est.0c06595. URL: https://doi.org/10.1021/acs.est.0c06595 (visited on 02/22/2024) (page 32).

[101] Hein Putter and Hans C. van Houwelingen. "Understanding Landmarking and Its Relation with Time-Dependent Cox Regression". en. In: *Statistics in Biosciences* 9.2 (Dec. 2017), pp. 489–503. ISSN: 1867-1772. DOI: 10.1007/s12561-016-9157-9. URL: https://doi.org/10.1007/s12561-016-9157-9 (visited on 02/29/2024) (page 32).

[102] Qian Di, Lingzhen Dai, Yun Wang, Antonella Zanobetti, Christine Choirat, Joel D. Schwartz, and Francesca Dominici. "Association of Short-term Exposure to Air Pollution With Mortality in Older Adults". In: *JAMA* 318.24 (Dec. 2017), pp. 2446–2456. ISSN: 0098-7484. DOI: 10.1001/jama.2017.17923. URL: https://doi.org/10.1001/jama.2017.17923 (visited on 08/29/2024) (page 33).

[103] Dan L. Crouse, Paul A. Peters, Perry Hystad, Jeffrey R. Brook, Aaron van Donkelaar, Randall V. Martin, Paul J. Villeneuve, Michael Jerrett, Mark S. Goldberg, C. Arden Pope, Michael Brauer, Robert D. Brook, Alain Robichaud, Richard Menard, and Richard T. Burnett. "Ambient PM2.5, O3, and NO2 Exposures and Associations with Mortality over 16 Years of Follow-Up in the Canadian Census Health and

Environment Cohort (CanCHEC)". In: *Environmental Health Perspectives* 123.11 (Nov. 2015). Publisher: Environmental Health Perspectives, pp. 1180–1186. DOI: 10.1289/ehp.1409276. URL: https://ehp.niehs.nih.gov/doi/10.1289/ehp.1409276 (visited on 07/29/2024) (page 33).

[104] D. L. Crouse, A. C. Erickson, T. Christidis, L. Pinault, A. van Donkelaar, C. Li, J. Meng, R. V. Martin, M. Tjepkema, P. Hystad, R. Burnett, A. Pappin, M. Brauer, and S. Weichenthal. "Evaluating the Sensitivity of PM2.5-Mortality Associations to the Spatial and Temporal Scale of Exposure Assessment". eng. In: *Epidemiology* 31.2 (Mar. 2020), pp. 168–176. ISSN: 1044-3983. DOI: 10.1097/ede.0000000000001136 (pages 33, 34).

[105] Jacob Lefler, Joshua Higbee, Richard Burnett, Majid Ezzati, Nathan Coleman, Dalton Mann, Julian Marshall, Matthew Bechle, Yuzhou Wang, Allen Robinson, and C. Pope. "Air pollution and mortality in a large, representative U.S. cohort: Multiple-pollutant analyses, and spatial and temporal decompositions". In: *Environmental Health* 18 (2019). DOI: 10.1186/s12940-019-0544-9 (pages 33, 34).

[106] D. Segersson, C. Johansson, and B. Forsberg. "Near-Source Risk Functions for Particulate Matter Are Critical When Assessing the Health Benefits of Local Abatement Strategies". eng. In: *Int J Environ Res Public Health* 18.13 (June 2021). Edition: 20210625. ISSN: 1661-7827 (Print) 1660-4601. DOI: 10.3390/ijerph18136847 (page 33).

[107] C. Arden Pope, S. Lefler Jacob, Ezzati Majid, D. Higbee Joshua, D. Marshall Julian, Kim Sun-Young, Bechle Matthew, S. Gilliat Kurtis, E. Vernon Spencer, L. Robinson Allen, and T. Burnett Richard.

"Mortality Risk and Fine Particulate Air Pollution in a Large, Representative Cohort of U.S. Adults". In: *Environmental Health Perspectives* 127.7 (2019), p. 077007. DOI: 10.1289/EHP4438 (pages 35, 135).

[108] George D. Thurston, Jiyoung Ahn, Kevin R. Cromar, Yongzhao Shao, Harmony R. Reynolds, Michael Jerrett, Chris C. Lim, Ryan Shanley, Yikyung Park, and Richard B. Hayes. "Ambient Particulate Matter Air Pollution Exposure and Mortality in the NIH-AARP Diet and Health Cohort". In: *Environmental Health Perspectives* 124.4 (Apr. 2016). Publisher: Environmental Health Perspectives, pp. 484–490. DOI: 10.1289/ehp.1509676. URL: https://ehp.niehs.nih.gov/doi/10.1289/ehp.1509676 (visited on 08/13/2024) (page 35).

[109] C. Arden Pope III, Richard T. Burnett, Michael J. Thun, Eugenia E. Calle, Daniel Krewski, Kazuhiko Ito, and George D. Thurston. "Lung Cancer, Cardiopulmonary Mortality, and Long-term Exposure to Fine Particulate Air Pollution". In: *JAMA* 287.9 (Mar. 2002), pp. 1132–1141. ISSN: 0098-7484. DOI: 10.1001/jama.287.9.1132. URL: https://doi.org/10.1001/jama.287.9.1132 (visited on 07/25/2024) (pages 35, 135).

[110] Lauren Pinault, Michael Tjepkema, Daniel L. Crouse, Scott Weichenthal, Aaron van Donkelaar, Randall V. Martin, Michael Brauer, Hong Chen, and Richard T. Burnett. "Risk estimates of mortality attributed to low concentrations of ambient fine particulate matter in the Canadian community health survey cohort". In: *Environmental Health* 15.1 (Feb. 2016), p. 18. ISSN: 1476-069X. DOI: 10.1186/s12940-

016-0111-6. URL: https://doi.org/10.1186/s12940-016-0111-6 (visited on 02/22/2024) (page 35).

[111] Jacopo Vanoli, Malcolm N. Mistry, Arturo De La Cruz Libardi, Pierre Masselot, Rochelle Schneider, Chris Fook Sheng Ng, Lina Madaniyazi, and Antonio Gasparrini. "Reconstructing individual-level exposures in cohort analyses of environmental risks: an example with the UK Biobank". en. In: *Journal of Exposure Science & Environmental Epidemiology* (Jan. 2024). Publisher: Nature Publishing Group, pp. 1–6. ISSN: 1559-064X. DOI: 10.1038/s41370-023-00635-w. URL: https://www.nature.com/articles/s41370-023-00635-w (visited on 01/15/2024) (page 40).

[112] Jeroen de Bont, Suganthi Jaganathan, Marcus Dahlquist, Åsa Persson, Massimo Stafoggia, and Petter Ljungman. "Ambient air pollution and cardiovascular diseases: An umbrella review of systematic reviews and meta-analyses". en. In: *Journal of Internal Medicine* 291.6 (2022). _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/joim.13467, pp. 779–800. ISSN: 1365-2796. DOI: 10.1111/joim.13467. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/joim.13467 (visited on 01/26/2024) (page 132).

[113] Hanne Krage Carlsen, Eva M. Andersson, Peter Molnár, Anna Oudin, Yiyi Xu, Janine Wichmann, Mårten Spanne, Emilie Stroh, Gunnar Engström, and Leo Stockfelt. "Incident cardiovascular disease and long-term exposure to source-specific air pollutants in a Swedish cohort". In: *Environmental Research* 209 (June 2022), p. 112698. ISSN: 0013-9351. DOI: 10.1016/j.envres.2022.112698. URL: https://

www.sciencedirect.com/science/article/pii/S0013935122000251 (visited on 01/29/2024) (page 133).

[114] Michael T. Osborne, Shady Abohashem, Nicki Naddaf, Taimur Abbasi, Hadil Zureigat, Kenechukwu Mezue, Ahmed Ghoneem, Tawseef Dar, Alexander J. Cardeiro, Nehal N. Mehta, Sanjay Rajagopalan, Zahi A. Fayad, and Ahmed Tawakol. "The combined effect of air and transportation noise pollution on atherosclerotic inflammation and risk of cardiovascular disease events". In: *Journal of Nuclear Cardiology* 30.2 (Apr. 2023), pp. 665–679. ISSN: 1071-3581. DOI: 10.1007/s12350-022-03003-7. URL: https://www.sciencedirect.com/science/article/pii/S1071358123001423 (visited on 01/26/2024) (page 133).

[115] Stacey E. Alexeeff, Kamala Deosaransingh, Stephen Van Den Eeden, Joel Schwartz, Noelle S. Liao, and Stephen Sidney. "Association of Long-term Exposure to Particulate Air Pollution With Cardiovascular Events in California". In: *JAMA Network Open* 6.2 (Feb. 2023), e230561. ISSN: 2574-3805. DOI: 10.1001/jamanetworkopen.2023.0561. URL: https://doi.org/10.1001/jamanetworkopen.2023.0561 (visited on 02/12/2024) (pages 133, 134).

[116] Wentao Zhu, Jiajie Cai, Yuchen Hu, Haodan Zhang, Xiao Han, Huiqiu Zheng, and Jing Wu. "Long-term exposure to fine particulate matter relates with incident myocardial infarction (MI) risks and post-MI mortality: A meta-analysis". In: *Chemosphere* 267 (Mar. 2021), p. 128903. ISSN: 0045-6535. DOI: 10.1016/j.chemosphere.2020.128903. URL: https://www.sciencedirect.com/science/article/pii/S0045653520331003 (visited on 02/01/2024) (page 134).

[117] "Associations of air pollution with obesity and body fat percentage, and modification by polygenic risk score for BMI in the UK Biobank". In: *Environmental Research* 185 (2020), p. 109364. ISSN: 0013-9351. DOI: 10.1016/j.envres.2020.109364. URL: https://www.sciencedirect.com/science/article/pii/S0013935120302577 (page 134).

[118] X. Wu, D. Braun, J. Schwartz, M. A. Kioumourtzoglou, and F. Dominici. "Evaluating the impact of long-term exposure to fine particulate matter on mortality among the elderly". In: *Science Advances* 6.29 (July 2020). Publisher: American Association for the Advancement of Science, eaba5692. DOI: 10.1126/sciadv.aba5692. URL: https://www.science.org/doi/full/10.1126/sciadv.aba5692 (visited on 08/29/2024) (pages 135, 141).

[119] Joel Schwartz, Qian Di, Weeberb Requia, Francesca Dominici, and Antonella Zanobetti. "A Direct Estimate of the Impact of PM2.5, NO2, and O3 Exposure on Life Expectancy Using Propensity Scores". In: *Epidemiology* 32 (2021), pp. 469–476. DOI: 10.1097/EDE.0000000000001354 (page 135).

[120] *Air Pollution in the UK 2022*. Tech. rep. DEFRA UK, Sept. 2022 (page 136).

[121] Elizabeth A. Richardson, Jamie Pearce, Helena Tunstall, Richard Mitchell, and Niamh K. Shortt. "Particulate air pollution and health inequalities: a Europe-wide ecological analysis". In: *International Journal of Health Geographics* 12.1 (July 2013), p. 34. ISSN: 1476-072X. DOI: 10.1186/1476-072X-12-34. URL: https://doi.org/10.1186/1476-072X-12-34 (visited on 06/20/2024) (page 136).

[122] Jules Kerckhoffs, Jibran Khan, Gerard Hoek, Zhendong Yuan, Ole Hertel, Matthias Ketzel, Steen Solvang Jensen, Fares Al Hasan, Kees Meliefste, and Roel Vermeulen. "Hyperlocal variation of nitrogen dioxide, black carbon, and ultrafine particles measured with Google Street View cars in Amsterdam and Copenhagen". In: *Environment International* 170 (Dec. 2022), p. 107575. ISSN: 0160-4120. DOI: 10.1016/j.envint.2022.107575. URL: https://www.sciencedirect.com/science/article/pii/S0160412022005025 (visited on 06/19/2024) (page 139).

[123] Arturo de la Cruz Libardi, Pierre Masselot, Rochelle Schneider, Emily Nightingale, Ai Milojevic, Jacopo Vanoli, Malcolm N. Mistry, and Antonio Gasparrini. "High resolution mapping of nitrogen dioxide and particulate matter in Great Britain (2003–2021) with multi-stage data reconstruction and ensemble machine learning methods". In: *Atmospheric Pollution Research* 15.11 (Nov. 2024), p. 102284. ISSN: 1309-1042. DOI: 10.1016/j.apr.2024.102284. URL: https://www.sciencedirect.com/science/article/pii/S1309104224002496 (visited on 08/21/2024) (pages 139, 149).

[124] Jochem O. Klompmaker, Nicole Janssen, Zorana J. Andersen, Richard Atkinson, Mariska Bauwelinck, Jie Chen, Kees de Hoogh, Danny Houthuijs, Klea Katsouyanni, Marten Marra, Bente Oftedal, Sophia Rodopoulou, Evangelia Samoli, Massimo Stafoggia, Maciej Strak, Wim Swart, Joost Wesseling, Danielle Vienneau, Bert Brunekreef, and Gerard Hoek. "Comparison of associations between mortality and air pollution exposure estimated with a hybrid, a land-use regression and a dispersion model". In: *Environment International* 146 (Jan. 2021), p. 106306.

ISSN: 0160-4120. DOI: 10.1016/j.envint.2020.106306. URL: https://www.sciencedirect.com/science/article/pii/S0160412020322613 (visited on 08/29/2024) (page 140).

[125] Zhou Jiang, Shuo Zhang, Tongyu Gao, Keying Chen, Yuxin Liu, Ying Liu, Ting Wang, and Ping Zeng. "Co-exposure to multiple air pollutants, genetic susceptibility, and the risk of myocardial infarction onset: a cohort analysis of the UK Biobank participants". In: *European Journal of Preventive Cardiology* (Dec. 2023), zwad384. ISSN: 2047-4873. DOI: 10.1093/eurjpc/zwad384. URL: https://doi.org/10.1093/eurjpc/zwad384 (visited on 01/30/2024) (page 140).

[126] X. Qiu, Y. Wei, Y. Wang, Q. Di, T. Sofer, Y. A. Awad, and J. Schwartz. "Inverse probability weighted distributed lag effects of short-term exposure to PM(2.5) and ozone on CVD hospitalizations in New England Medicare participants - Exploring the causal effects". eng. In: *Environ Res* 182 (Mar. 2020). Edition: 20191230, p. 109095. ISSN: 0013-9351 (Print) 0013-9351. DOI: 10.1016/j.envres.2019.109095 (page 141).

[127] Chen Chen, Hong Chen, Aaron van Donkelaar, Richard T. Burnett, Randall V. Martin, Li Chen, Michael Tjepkema, Megan Kirby-McGregor, Yi Li, Jay S. Kaufman, and Tarik Benmarhnia. "Using Parametric g-Computation to Estimate the Effect of Long-Term Exposure to Air Pollution on Mortality Risk and Simulate the Benefits of Hypothetical Policies: The Canadian Community Health Survey Cohort (2005 to 2015)". In: *Environmental Health Perspectives* 131.3 (Mar. 2023). Publisher: Environmental Health Perspectives, p. 037010. DOI: 10.

1289/EHP11095. URL: https://ehp.niehs.nih.gov/doi/full/10.1289/EHP11095 (visited on 07/24/2024) (page 141).

[128] Rina So, Zorana J. Andersen, Jie Chen, Massimo Stafoggia, Kees de Hoogh, Klea Katsouyanni, Danielle Vienneau, Sophia Rodopoulou, Evangelia Samoli, Youn-Hee Lim, Jeanette T. Jørgensen, Heresh Amini, Tom Cole-Hunter, Seyed Mahmood Taghavi Shahri, Matija Maric, Marie Bergmann, Shuo Liu, Shadi Azam, Steffen Loft, Rudi G. J. Westendorp, Laust H. Mortensen, Mariska Bauwelinck, Jochem O. Klompmaker, Richard Atkinson, Nicole A. H. Janssen, Bente Oftedal, Matteo Renzi, Francesco Forastiere, Maciek Strak, Lau C. Thygesen, Bert Brunekreef, Gerard Hoek, and Amar J. Mehta. "Long-term exposure to air pollution and mortality in a Danish nationwide administrative cohort study: Beyond mortality from cardiopulmonary disease and lung cancer". en. In: *Environment International* 164 (June 2022), p. 107241. ISSN: 0160-4120. DOI: 10.1016/j.envint.2022.107241. URL: https://www.sciencedirect.com/science/article/pii/S0160412022001672 (visited on 04/05/2023) (page 142).

[129] Larry Goldstein and Bryan Langholz. "Asymptotic Theory for Nested Case-Control Sampling in the Cox Regression Model". In: *The Annals of Statistics* 20.4 (1992). Publisher: Institute of Mathematical Statistics, pp. 1903–1928. ISSN: 0090-5364. URL: https://www.jstor.org/stable/2242373 (visited on 04/14/2023) (page 147).

[130] Vidal Essebag, Robert W. Platt, Michal Abrahamowicz, and Louise Pilote. "Comparison of nested case-control and survival analysis methodologies for analysis of time-dependent exposure". In: *BMC Medical Research Methodology* 5.1 (Jan. 2005), p. 5. ISSN: 1471-2288. DOI:

10.1186/1471-2288-5-5. URL: https://doi.org/10.1186/1471-2288-5-5 (page 148).