Environment International 147 (2021) 106371

Contents lists available at ScienceDirect

Environment International

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

Modeling multi-level survival data in multi-center epidemiological cohort studies: Applications from the ELAPSE project

Evangelia Samoli [a,*], Sophia Rodopoulou [a], Ulla A. Hvidtfeldt [b], Kathrin Wolf [c], Massimo Stafoggia [d,e], Bert Brunekreef [f], Maciej Strak [f,g], Jie Chen [f], Zorana J. Andersen [h], Richard Atkinson [i], Mariska Bauwelinck [j], Tom Bellander [e,k], Jørgen Brandt [l], Giulia Cesaroni [d], Francesco Forastiere [m], Daniela Fecht [n], John Gulliver [o], Ole Hertel [l], Barbara Hoffmann [p], Kees de Hoogh [q,r], Nicole A.H. Janssen [f], Matthias Ketzel [l,s], Jochem O. Klompmaker [f,g], Shuo Liu [h], Petter Ljungman [e,t], Gabriele Nagel [u], Bente Oftedal [v], Göran Pershagen [e,k], Annette Peters [c], Ole Raaschou-Nielsen [b,l], Matteo Renzi [d], Doris T. Kristoffersen [w], Gianluca Severi [d], Torben Sigsgaard [x], Danielle Vienneau [q,r], Gudrun Weinmayr [u], Gerard Hoek [f], Klea Katsouyanni [a,m]

[a] Dept. of Hygiene, Epidemiology and Medical Statistics, Medical School, National and Kapodistrian University of Athens, 75 Mikras Asias Str, 115 27 Athens, Greece
[b] Danish Cancer Society Research Centre, Copenhagen, Denmark
[c] Institute of Epidemiology, Helmholtz Zentrum München, Neuherberg, Germany
[d] Department of Epidemiology, Lazio Region Health Service ASL Roma 1, Rome, Italy
[e] Institute of Environmental Medicine, Karolinska Institutet, Stockholm, Sweden
[f] Institute for Risk Assessment Sciences (IRAS), Utrecht University, Postbus 80125, 3508 TC Utrecht, the Netherlands
[g] National Institute for Public Health and the Environment (RIVM), Antonie van Leeuwenhoeklaan 9, 3721 MA Bilthoven, the Netherlands
[h] University of Copenhagen, Department of Public Health, Section of Environmental Health, Øster Farimagsgade 5, 1014, Copenhagen, Denmark
[i] Population Health Research Institute, St George's, University of London, Cranmer Terrace, London SW17 0RE, UK
[j] Interface Demography, Department of Sociology, Vrije Universiteit Brussel, Brussels, Belgium
[k] Centre for Occupational and Environmental Medicine, Region Stockholm, Stockholm, Sweden
[l] Department of Environmental Science, Aarhus University, Frederiksborgvej 399, Roskilde, Denmark
[m] NIHR HPRU Health Impact of Environmental Hazards, Environmental Research Group, Analytical, Environmental & Forensic Sciences, King's College London, UK
[n] Small Area Health Statistics Unit, MRC Centre for Environment and Health, School of Public Health, Imperial College London, Norfolk Place, London W2 1PG, UK
[o] Centre for Environmental Health and Sustainability & School of Geography, Geology and the Environment, University of Leicester, Leicester, UK
[p] Institute for Occupational, Social and Environmental Medicine, Medical Faculty, Heinrich-Heine-University of Düsseldorf, Germany
[q] Swiss Tropical and Public Health Institute, Basel, Switzerland
[r] University of Basel, Basel, Switzerland
[s] Global Centre for Clean Air Research (GCARE), University of Surrey, Guildford GU2 7XH, UK
[t] Department of Cardiology, Danderyd University Hospital, Stockholm, Sweden
[u] Institute of Epidemiology and Medical Biometry, Ulm University, Ulm, Germany
[v] Department of Environmental Health, Norwegian Institute of Public Health, Oslo, Norway
[w] Cluster for Health Services Research, Norwegian Institute of Public Health, Oslo, Norway
[x] Department of Public Health, Environment Occupation and Health, Danish Ramazzini Centre, Aarhus University, Aarhus, Denmark

ARTICLE INFO

Handling Editor: Shoji F. Nakayama

Keywords:
Air pollution
Cox model
Frailty models
Health effects

ABSTRACT

Background: We evaluated methods for the analysis of multi-level survival data using a pooled dataset of 14 cohorts participating in the ELAPSE project investigating associations between residential exposure to low levels of air pollution ($PM_{2.5}$ and NO_2) and health (natural-cause mortality and cerebrovascular, coronary and lung cancer incidence).

Methods: We applied five approaches in a multivariable Cox model to account for the first level of clustering corresponding to cohort specification: (1) not accounting for the cohort or using (2) indicator variables, (3) strata, (4) a frailty term in frailty Cox models, (5) a random intercept under a mixed Cox, for cohort

* Corresponding author at: Department of Hygiene and Epidemiology, National and Kapodistrian University of Athens Medical School, 75 Mikras Asias Street, 115 27 Athens, Greece.

https://doi.org/10.1016/j.envint.2020.106371
Received 30 October 2020; Received in revised form 18 December 2020; Accepted 25 December 2020
Available online 12 January 2021
0160-4120/© 2021 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

Mixed models
Multi-level analysis

identification. We accounted for the second level of clustering due to common characteristics in the residential area by (1) a random intercept per small area or (2) applying variance correction. We assessed the stratified, frailty and mixed Cox approach through simulations under different scenarios for heterogeneity in the underlying hazards and the air pollution effects.

Results: Effect estimates were stable under approaches used to adjust for cohort but substantially differed when no adjustment was applied. Further adjustment for the small area grouping increased the effect estimates' standard errors. Simulations confirmed identical results between the stratified and frailty models. In ELAPSE we selected a stratified multivariable Cox model to account for between-cohort heterogeneity without adjustment for small area level, due to the small number of subjects and events in the latter.

Conclusions: Our study supports the need to account for between-cohort heterogeneity in multi-center collaborations using pooled individual level data.

1. Introduction

Multi-center collaborations provide an opportunity for harmonized epidemiological analyses in large data sets that increase statistical power to detect associations and provide more precise estimates over a larger range of exposures. In general, two types of methods have been used in the majority of multi-center studies. The preferred method is primarily driven by data protection policies within each center that impact the feasibility of sharing individual participant data. One method is to follow a prospective meta-analysis design (Beelen et al., 2014; Katsouyanni et al., 2001) where each collaborating center performs the analysis of their data following a harmonized statistical protocol and no data transfer occurs. Subsequently, meta-data (effect estimates with their corresponding standard errors) are shared and pooled together through meta-analysis techniques. A second method is pooled data analysis where the data are pooled into one data set and then analyzed together (Lindsay et al., 2019; Pedersen et al., 2013). Each method has its advantages and disadvantages, largely depending also on the design of the studies that contribute data. An advantage of meta-analysis is the opportunity of the investigation of between studies effects' heterogeneity, while pooled analysis enhances statistical power. In theory both methods provide identical overall effect estimates under specific assumptions following a fixed or random effects model (that account for within studies' or additionally for between studies' variation correspondingly) (Basagaña et al., 2018). In practice, this equivalence mainly applies to randomized clinical trials, where confounding issues, as a main source of potential bias, are controlled by design, in contrast to observational studies where confounding control usually differs between participating studies depending on data availability. Further different exposure or outcome measurement errors and various related forms of bias, as information or selection bias due to loss to follow-up, could also lead to different results, while the potential difference in effect of between and within studies exposure contrast may result in differences between the fixed and random effect model (Basagaña et al., 2018).

The analysis of pooled multilevel survival data presents specific challenges due to the differences in underlying hazards between studies. Previous comparisons between available methods focused on specific approaches (eg. between specific conditional models) and have been largely compared within randomized clinical trials settings (Austin, 2017; Glidden and Vittinghoff, 2004).

The European Study of Cohorts for Air Pollution Effects (ESCAPE) investigated air pollution health effects using meta-analysis to data from 22 existing European cohorts (Beelen et al., 2014; Raaschou-Nielsen et al., 2013; Stafoggia et al., 2014; Cesaroni et al., 2014). In the present study, within the context of the "Effects of Low-Level Air Pollution: A Study in Europe (ELAPSE, www.elapseproject.eu)" project, 13 of the ESCAPE cohorts and one additional cohort agreed to pool data in order to investigate the association between long-term residential exposure to low levels of specific air pollutants and several health outcomes. Considering the generally small health related relative risks in relation to air pollution, the initial idea was that the pooled data would provide improved exposure contrasts between cohorts and increased statistical power compared to the meta-analysis approach. The latter is particularly important considering that the main research question in ELAPSE was the investigation of health effects following exposure to low-levels of air pollution, hence the pooled data provide larger sample sizes in the left extreme of the distributions, where our interest lies. Further pooled analysis is a more straightforward approach providing a single estimate of the concentration –response function instead of combining uncertain estimates obtained from moderately small cohorts. In the process of finalizing the methodology for the analysis of the pooled cohorts' data set, we assessed several approaches to account for the first level of grouping in the data, which is the underlying participating cohort studies. In addition, since the focus of the ELAPSE project is on the association between long-term residential exposure to air pollutants and health, the integration of a second level of clustering in the data that corresponds to the small area level of the participants' residence is inherent, as residents in the same area may share common social and exposure characteristics that may not be fully captured by the individual and area-level confounders included in our models. This multi-level structure in our data makes the decision of the modeling choice challenging, as not accounting for it may deflate the confidence interval, while the opposite may not exploit between-cohort exposure contrast.

In the present paper we report the results from the comparison of different methodological approaches for the analysis of multi-level survival data in pooled epidemiological cohorts. Our extensive sensitivity analysis within the ELAPSE project intended to identify the optimal approach to account for the different underlying original cohorts and the multi-level clustering of such data in our pooled database.

2. Data and methods

2.1. Data

Eight European prospective studies comprising 14 individual cohorts contributed data for the present analyses on 325,367 participants enrolled in different periods ranging from 1992 to 2005 (in alphabetical order): Cardiovascular Effects of Air Pollution and Noise in Stockholm (CEANS in Stockholm, Sweden consisting of four separate cohorts, namely the Screening Across the Lifespan Twin Study (SALT), the Stockholm Diabetes Preventive Program (SDPP), the cohort study of 60 year olds (SIXTY) and the Swedish National Study of Aging and Care in Kungsholmen (SNAC-K)); Diet, Cancer and Health (DCH in Copenhagen and Aarhus, Denmark); Danish Nurse Cohort (DNC nationwide Denmark, with two cohorts in 1993 and 1999); European Prospective Investigation into Cancer and Nutrition, the Netherlands (EPIC-NL in four cities, The Netherlands with two cohorts, Morgen and Prospect); Etude Epidémiologique auprès de femmes de la Mutuelle Générale de l'Education Nationale (E3N, nationwide France); Heinz Nixdorf Recall study (HNR in the Ruhr area, Germany); Kooperative Gesundheits forschung in der Region Augsburg (KORA in Augsburg, Germany with two cohorts, S3 and S4); and Vorarlberg Health Monitoring and Prevention Programme (VHM&PP in Vorarlberg, Austria). Supplementary Table S1 presents the cohorts' main characteristics.

In this paper, we focus on natural-cause mortality (International

Classification of Diseases (ICD-9: 001-779; ICD-10: A00-R99), incidence of lung cancer (ICD-9: 162.2–162.9; ICD-10:C34.0-C34.9) and incidence of coronary and cerebrovascular events. Coronary events were defined as hospitalizations with principal diagnosis of acute myocardial infarction or other acute and sub-acute forms of ischemic heart disease (ICD-9: 410, 411, 427.5; ICD-10: I20.0, I21, I22, I23, I46) and out-of-hospital deaths from ischemic heart diseases (ICD-9: 410–414, 427.5; ICD-10: I20-I25, I46). Cerebrovascular events were defined as hospitalizations with principal diagnosis of ischemic stroke (ICD-9: 433.x1, 434; ICD-10: I63), hemorrhagic stroke (ICD-9: 431; ICD-10: I61) or unspecified stroke (ICD-9: 436; ICD-10: I64) and out-of-hospital deaths from cerebrovascular diseases (ICD-9: 431–436; ICD-10: I61-I64). Information on covariates have been collected at baseline including age, sex, body mass index, smoking (with varying degree of detail between cohorts), employment and marital status, educational level, and diet (EPIC cohorts only), as well as a limited number of socio-economic status area level variables available for census years including mean income. Details on the exposure assessment have been previously reported (de Hoogh et al., 2018). In short, concentrations of particulate matter with diameter < 2.5 μm ($PM_{2.5}$) and nitrogen dioxide (NO_2), ozone (O_3) and black carbon (BC) were estimated using Europe-wide land use regression models at 100 × 100 m grids, which incorporated AirBase routine monitoring data for $PM_{2.5}$, NO_2 and O_3, and ESCAPE monitoring data for BC as dependent variables and land use, traffic data, satellite observations and dispersion model output as predictor variables. We then assigned to the baseline residential address of the cohort participants the estimated concentration in the 100 × 100 m grid the address was located. Here we focus on $PM_{2.5}$ and NO_2 as an example of two pollutants with spatial distributions that present distinct patterns between them and vary between cohorts (Brunekreef et al., 2021).

We defined our confounder model based upon the ESCAPE study (Beelen et al., 2014). One main challenge in pooling data previously collected by different cohort studies is the availability of the optimal set of covariates across the different studies that will be used in the common statistical approach. Even when similar covariates are available, incompatibility in covariate coding between cohorts may inhibit harmonization and use in a pooled analysis. Within ELAPSE we applied several options for harmonizing covariates considered to be important in the association between long-term residential exposure to specific air pollutants and several health outcomes, such as smoking and diet variables, and assessed the sensitivity of the main analysis to the inclusion of a set of variables only available in a subset of the initial cohorts. Further, multiple imputation techniques have been previously proposed for the imputation of variables completely missing from one study based on the available ones (Resche-Rigon et al., 2013; Jolani et al., 2015), but we decided against this approach considering the observational design of our studies and their possible underlying population differences in lifestyle related characteristics. Instead, we opted for extensive sensitivity analyses regarding the inclusion of specific cohorts/variables in the model. Extensive sensitivity analyses regarding covariate harmonization (which is not the focus of this paper and is described in the ELAPSE report (Brunekreef et al., 2021) indicated that that after a main model was defined with covariates available for all cohorts, mainly the inclusion or exclusion of cohorts was driving the effect estimates more than the inclusion of extra covariate control not available for all cohorts. Hence we decided to include all cohorts in the main model adjusting for a set of harmonized covariates, available in all cohorts, as described in the Methods section below.

2.2. Methods

We assessed five approaches to account for the first level of clustering in our pooled dataset corresponding to cohort specification (Debray et al., 2015; Therneau and Grambsch, 2000): (1) not accounting for the cohort or (2) including a categorical covariate in the model with indicator variables to identify cohorts, (3) strata per cohort, (4) a frailty term for cohort identification (as applied in frailty Cox models that are random effects models for survival data accounting for clustering by a random ("frailty") intercept following (usually) a gamma distribution), (5) a random intercept per cohort under a mixed Cox model (mixed models allowing for both fixed and random effects (following a Gaussian distribution) and accounting for clustering in various levels). We further accounted for the second level of clustering at the small area level by either (1) a random intercept per small area or (2) applying variance correction, recognizing that these allow for different interpretations following either a conditional or marginal approach correspondingly.

We finally assessed the stratified, frailty and mixed Cox approach for cohort adjustment through 1 000 simulations under nine scenarios for varying amount of heterogeneity in baseline hazards and pollutants' effects between cohorts. The choice of these three methods was based on the violation of the proportional hazards (PH) assumption that was anticipated in our data and due to their differences in accounting for exposure contrast.

We applied Cox proportional hazards models using age as the time axis given the evidence that this results in better adjustment for potential confounding by age (Thiébaut and Bénichou, 2004). Censoring occurred at the time of the event of interest, death from causes other than natural for analysis of natural mortality, other cancer incidence for analysis of lung cancer incidence or other CVD incidence for coronary and cerebrovascular events incidence, emigration, loss to follow-up for other reasons (that ranged from 1% to 15%, with larger percentages in the HNR and KORA cohorts that contributed <4% in the pooled data), or at the end of follow-up, whichever came first. We stratified the baseline hazard function by sex and controlled for harmonized common variables that were included in the main analysis: smoking status at baseline (present, former vs never), smoking intensity among current smokers (continuous in cigarettes per day as a linear and quadratic term), smoking duration (continuous in years), body mass index at baseline (as a four-level variable following the World Health Organization (WHO) categorization as underweight: <18.5 kg/m^2, normal weight: 18.5 kg/m^2–24.9 kg/m^2, overweight: 25.0 kg/m^2–29.9 kg/m^2, obese: ≥30.0 kg/m^2), marital status at baseline (as a four-level variable for single, married/living with partner, divorced/separated and widowed), employment status at baseline (employed vs other) and mean income at small area level in 2001 (continuous in euros). In coronary and cerebrovascular events analysis we further adjusted for educational level (as a three-level categorical variable, indicating low, high school and higher education) and smoking intensity and duration among former smokers (continuous in cigarettes per day as a linear and quadratic term), as the information was available in all the cohorts contributing these health data. Mean income was the only area level variable available across participating cohorts, while the definition of small area was the municipality for most participating cohorts. We also compared the approaches considered to account for the two levels of clustering in our pooled data using Cox models not including mean income in the natural mortality analysis, in order to assess the robustness of our findings.

The pooled cohort included a total of 1310 small areas, with a median number of subjects of 15 per area and a median number of two deaths in the full follow-up period. It is important to note, that the number of small areas varied greatly between cohorts (from 26 areas in CEANS-SNACK to 711 in CEANS-SALT), as also did the number of participants, even in small areas within the same cohort (the most extreme example being EPIC-NL-Morgen with a range between 1 and >1400 subjects between areas related to the neighborhood being rural or highly urban in Amsterdam).

We assessed the PH assumption using log–log plots and compared models in terms of the Akaike Information Criteria (AIC).

2.2.1. Approaches in a snapshot

Not accounting for the original cohort in the model (approach 1) assumes that baseline hazards and hazard ratios (HRs) are the same for each cohort after including covariates, hence this was labeled as a naive

approach. This approach has the advantage of exploiting the full exposure contrast across cohorts. In terms of estimation, it is an unconditional model applying a maximum partial likelihood (ML), generally using the Newton–Raphson algorithm (NR). When we include indicator variables to characterize the cohorts (approach 2) the model follows a fixed effects approach. It is a conditional model assuming cohorts act proportionally on the baseline hazard risk, while estimation again follows from NR. Under this approach, only within-study exposure contrasts are exploited, as indicator variables adjust for differences between cohorts. Using a stratified model for the contributing cohort (approach 3) allows baseline hazards to vary by cohort and relaxes PH assumption between cohorts. Under stratification, the baseline hazards play no role in the estimation (again by the ML approach) and the model does not estimate between-cohort variance. The lack of structure makes this choice the most general among the conditional models, although, as it is conditional, there are no between-cohorts comparisons and all information comes from the within-cohort comparisons. This disadvantage is partly overcome by large sample sizes as then the Cox model's fit is equivalent to a frailty one with a gamma distribution (O'Quigley and Stare, 2002; Hosmer et al., 2008). The remaining two approaches (frailty (approach 4) and mixed Cox (approach 5); (Hosmer et al., 2008) follow a random effects approach with random intercepts for participating cohorts. They estimate between-cohort heterogeneity but assume different underlying distributions for the baselines hazards. The frailty model treats cohort effects as a sample from a (usually) gamma distribution, where the frailties represent unmeasured factors affecting cohort-specific baseline hazard risks and are assumed to act multiplicatively on the average baseline hazard. Estimation is typically made through a penalized partial likelihood algorithm. Frailty models are a special case of the mixed effects survival models (Austin, 2017). A mixed Cox model with random intercept for the underlying cohorts is basically the same as the frailty, but it assumes a Gaussian distribution of underlying hazards. Several authors have described methods for assessing whether a frailty model in which the random effects follow a gamma distribution fit the data well (Austin, 2017). However, they did not compare between different distributions, but only assessed whether the gamma distribution is reasonable. These latter conditional approaches exploit within- and between-cohort exposure contrasts.

We further assessed two methods for adjusting for the second level of clustering inherent in our data due to the subjects' residence in the same small area: (1) the mixed Cox model presented above expanded to account for this second level of correlated observations with a random intercept at small area level and (2) a variance correction approach in the model with no control, indicator variables or strata per cohort to correct the standard errors of our effect estimates for this correlation. The variance correction applies a robust or sandwich-type variance estimator to account for the clustering of subjects (Austin, 2017; Therneau and Grambsch, 2000). As detailed above, the mixed Cox model follows a conditional modeling approach with cluster-specific interpretation while the variance correction follows a marginal approach with a population-average interpretation.

All analyses were done in R version 3.4.0 with packages: *survival*, *coxme*, *Matrix*, *foreach*, *multcomp*, *survey*, *Hmisc*, *ggplot2*, *frailtySurv*, *survsim*, *eha*, *stamod*. R provides several options for frailty and Cox mixed models. Specifically, the function *frailty* of the *survival* package allows for either: 1) a Gamma distribution and a choice between estimation-maximization algorithm (that is the default and was used in our analysis) or restricted maximum –likelihood (REML); or 2) a Gaussian frailty under REML estimation. The R package *coxme* applies a Gaussian Cox mixed effects models under EM algorithm. We provide extracts of our code to fit the approaches in the Supplementary material.

2.2.2. Simulations set-up

We further assessed the stratified, frailty and mixed Cox approach through simulations under different scenarios for the heterogeneity both in the underlying hazards and the pollution effects captured by a random slope in the exposure. Finally, we wanted to assess the impact of not accounting for a random slope in our small effect estimates in the presence of underlying heterogeneity. Up to date, previous comparisons (O'Quigley and Stare, 2002; Hosmer et al., 2008; Giganti et al., 2015) have been based in clinical settings or associations with larger estimated effects as compared to the small effects encountered in environmental epidemiology.

We used the *genfrail* function of *frailtySurv* R package to simulate clustered survival data with a frailty and different exposure effects per cohort to investigate the association between $PM_{2.5}$ exposure and natural-cause mortality. We considered eight clusters, i.e. eight cohorts, with sample sizes derived from a uniform distribution with lower and upper bound 5 000 and 120 000 persons, respectively, based on our cohorts' population sizes. The data were generated assuming different beta per cohort for the association between pollution and mortality and a normal distribution for the $PM_{2.5}$ concentrations, with mean 14.67 μg/m^3 and standard deviation 2.01 μg/m^3 (Beelen et al., 2014). We simulated data for different amounts of heterogeneity between cohorts and between pollutant effects. The true effect estimates were sampled from a normal distribution with mean 0.007 and three alternative variance values: $1.05 \cdot 10^{-6}$; $2.11 \cdot 10^{-6}$ and $4.21 \cdot 10^{-6}$, to account for small, medium and large heterogeneity in air pollution effects between cohorts, respectively. These parameters represent the log(HR) and the variance expressed per 1 μg/m^3 $PM_{2.5}$. These values were based on a meta-analysis within the Shaping EUROpean policies to promote HEALTH equity (EUROHEALTHY) project (http://www.euro-healthy.eu, Deliverable D. 3.1). Frailty values were sampled from a Gamma (0.5, 0.5), a Gamma (1, 1) and a Gamma (2, 2) distribution to consider small, medium and large heterogeneity between underlying hazards in the cohorts. Hence, nine scenarios resulted from the combination of the degree of frailty and the heterogeneity of effects between cohorts.

We considered an exponential distribution for the survival time with baseline hazard λ equal to 0.017, following Bender et al. (Bender et al., 2005), who relate the expected value of the survival time to the tables for life expectancy. In detail, for survival time T, $E(T) = 1/\lambda$ for the Exponential distribution, while the mean life expectancy (E(T)) for age 20 is about 60 years, so $1/\lambda = 60 => \lambda = 1/60 = 0.017$. We sampled follow up time from a uniform distribution (0, 15) to simulate non-informative right censorship and an event rate of about 10%. In an extra step we set the follow up time at 15 years for all censored observations to obtain no loss to follow up in the simulated survival data, as the association under investigation is for mortality hence censoring is only due to the observation time.

We compared the following Cox models under the nine heterogeneity scenarios: a) a stratified Cox model by cohort, b) a frailty Cox with a random intercept per cohort (frailty Cox), c) a stratified Cox model with a random slope for the pollutant's effect and d) a mixed Cox model with random intercept and slope.

We generated 1000 simulated data sets and evaluated the performance of the Cox models in terms of bias (mean difference between true and estimated effect estimate), root mean square error (MSE) and coverage probabilities (% of simulations where the 95% confidence interval (CI) contains the true effect).

3. Results

3.1. ELAPSE data analysis

The sample size of the pooled database that differed according to the outcome, as not all cohorts contributed incidence data (Table 1). Although the sample size for the analysis of the lung cancer incidence was almost double to that for coronary events, the number of cases for the latter were >10 000 as compared to about 4 000 incident lung cancer cases. Mean levels of $PM_{2.5}$ were 15 μg/m^3 (standard deviation (SD) 3.2 μg/m^3) in natural mortality and lung cancer analysis and slightly lower in the CVD outcomes analysis (related to fewer cohorts in the CVD

Table 1
Description of pooled database per analyzed outcome.

Variable	Natural mortality	Incidence of		
		Coronary events	Stroke	Lung cancer
Participating cohorts	All	All except VHM&PP and E3N	All except VHM&PP and E3N	All except KORA
Participants with complete covariate data	325 367	137 175	137 175	307 550
Person Years of Follow up	6 339 553	2 167 286	2 167 286	5 561 379
Number of events	47 117	10 037	9261	3956
Baseline age, years (mean, (standard deviation))	49 (13)	54 (9)	54 (9)	48 (13)

Note: Refer to Supplemental Table S1 for a description of the individual cohorts. E3N: Etude Epidémiologique auprès de femmes de la Mutuelle Générale de l'Education Nationale; KORA: Kooperative Gesundheits forschung in der Region Augsburg; VHM&PP: Vorarlberg Health Monitoring and Prevention Programme.

analysis). Mean NO_2 levels were 25 μg/m^3 (SD 8 μg/m^3) in natural mortality and lung cancer analysis and slightly higher in the CVD outcomes analysis. In all analyses the lowest levels were observed in CEANS-SDPP ($PM_{2.5}$ 7.5 μg/m^3 and NO_2 15.4 μg/m^3) and the highest in HNR ($PM_{2.5}$ 19.5 μg/m^3 and NO_2 37.8 μg/m^3) (Brunekreef et al., 2021).

All methods that applied adjustment for cohorts resulted in comparable effect estimates irrespective of the pollutant or the outcome. Fig. 1 and Supplementary Table S2 present the results from the five approaches for cohort adjustment in terms of HRs and 95% CI for natural-cause mortality associated with a 5 μg/m^3 increase in $PM_{2.5}$ or 10 μg/m^3 increase in NO_2. Corresponding results from incidence data are presented in Supplementary Tables S3 and S4. In comparison to not controlling for cohort, the effect estimates for all methods were increased for natural mortality for both pollutants. Results were more variable for incidence analyses, as effect estimates increased for NO_2 and slightly decreased for $PM_{2.5}$. The CIs became wider with cohort adjustment for all endpoints, as expected due to the decrease in the exploited exposure contrast. AIC indicated the stratification approach as the optimal fit for all outcome-pollutant pairs (Supplementary Tables S2–S4). The deviation from the PH assumption in the underlying pooled data (as presented by crossing hazards in Supplementary Fig. S1) supports the choice to account for the cohort of origin in the main analysis, using a stratification approach.

Table 2 presents effect estimates for natural-cause mortality associated with the corresponding increase in $PM_{2.5}$ or NO_2 additionally controlling for the small area level grouping of the data. Corresponding results from incidence data are presented in Supplementary Tables S5 and S6. As three large cohorts (DNC, VHM&PP and E3N) did not provide identification codes for the small area the pooled databases applied for this analysis differ from those in Table 1. Hence, changes in the range of the estimates compared to results in Fig. 1, both upwards (e.g. most approaches natural-cause mortality and $PM_{2.5}$) and downwards (lung cancer and $PM_{2.5}$) are attributed to the different cohorts analyzed. In

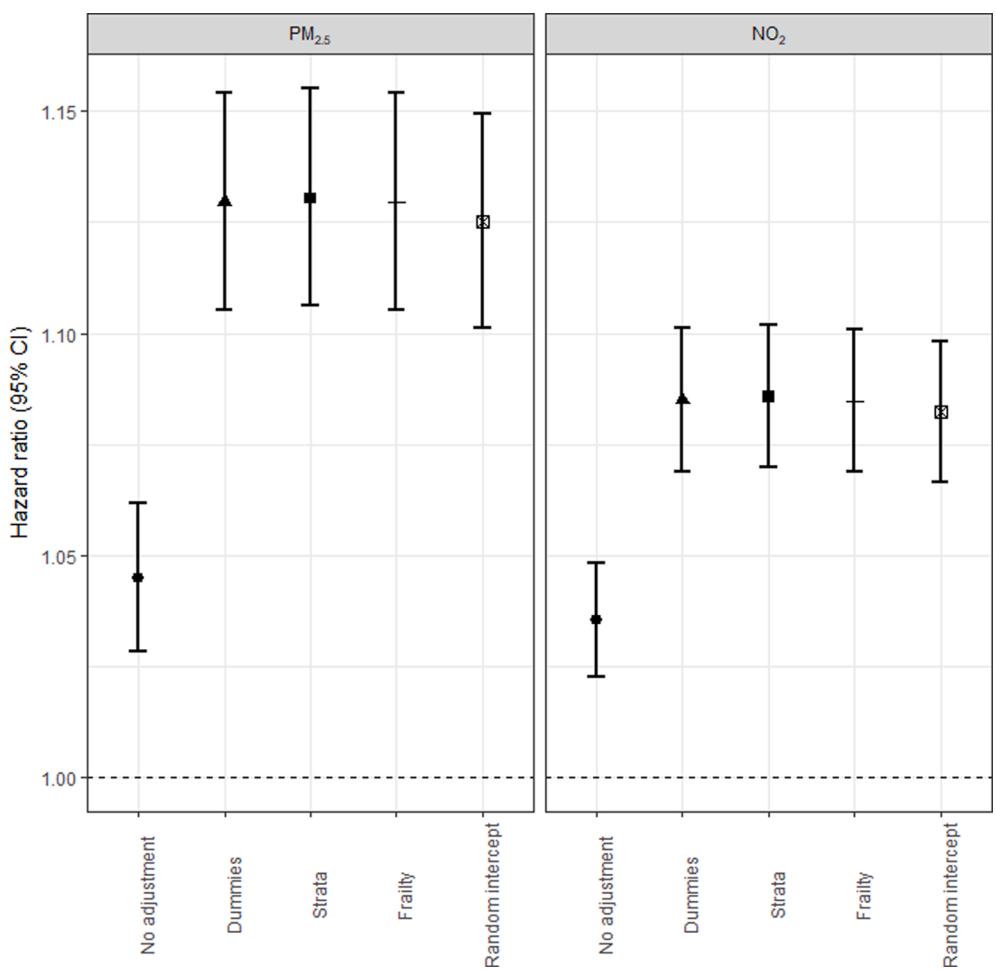

Fig. 1. Hazard ratios (and associated 95% confidence intervals (CI)) for natural-cause mortality associated with a 5 μg/m^3 increase in $PM_{2.5}$ and 10 μg/m^3 increase in NO_2 using several approaches for cohort adjustment. All models use age as the time axis, stratify by sex and adjust for body mass index, employment and marital status, smoking status, smoking intensity and duration and mean small area income.

Table 2
Hazard ratios ((HR) and associated 95% confidence intervals (CI)) for natural-cause mortality associated with a 5 μg/m^3 increase in $PM_{2.5}$ or 10 μg/m^3 increase in NO_2 using several approaches for adjustment for cohort and small area (SA) level in the subset of cohorts with indication code of small area available. All models use age as the time axis, stratify by sex and control for body mass index, employment and marital status, smoking status, smoking intensity and duration and mean SA income. All cohorts except DNC, VHM&PP and E3N, n = 116 807 subjects.[a]

Modeling approach	$PM_{2.5}$ HR (95%CI)	NO_2 HR (95%CI)
No control for cohort		
No control for SA	1.006 (0.980, 1.033)	1.035 (1.014, 1.055)
Variance correction for SA	1.006 (0.961, 1.053)	1.035 (0.983, 1.088)
Indicator for cohort		
No control for SA	1.191 (1.122, 1.264)	1.080 (1.054, 1.107)
Variance correction for SA	1.191 (1.062, 1.336)	1.080 (1.043, 1.119)
Strata for cohort		
No control for SA	1.182 (1.113, 1.255)	1.075 (1.049, 1.102)
Variance correction for SA	1.182 (1.048, 1.333)	1.075 (1.036, 1.117)
Mixed Cox		
No control for SA	1.180 (1.113, 1.250)	1.079 (1.053, 1.106)
Two levels adjustment for cohort and SA	1.090 (1.020, 1.165)	1.042 (1.012, 1.074)

[a] Models with smaller sample size due to missing values for small area characterization in DNC, VHM&PP, E3N; n in Table S2 = 325 637; n in Table 2 = 116 807. DNC: Danish Nurses Cohort; E3N: Etude Epidémiologique auprès de femmes de la Mutuelle Générale de l'Education Nationale; VHM&PP: Vorarlberg Health Monitoring and Prevention Programme.

order to assess the impact of the adjustment at small area level we present results restricted to cohorts that provided the relevant identification and hence comparisons are for within-cohort adjustment methods. Accounting for the small area grouping of the data through variance correction methods, moderately (coronary events) or substantially (natural-cause mortality and $PM_{2.5}$) increased the CIs, depending on the association. The impact differed both by outcome but also by pollutant reflecting the differences in the small number of cases at this level and the exposure contrast exploited. Applying a two- level mixed Cox model resulted in slightly smaller effect estimates and wider CIs. Comparison of these approaches in natural mortality models excluding mean income as an area level covariate followed the same pattern between methods as in the case of its inclusion. Effect estimates with (Table 2) vs without (Supplementary Table S7) mean income were slightly higher in the $PM_{2.5}$ analysis and lower in the NO_2 analysis (except when not accounting for cohort).

3.2. Simulation results

Table 3 presents the simulation results for the comparison of modeling approaches under the nine scenarios for heterogeneity between cohorts in baseline hazard and pollutant's effect. All approaches provided identical results in terms of bias, MSE or coverage probabilities under most scenarios. Indicatively the stratified vs the frailty model resulted in practically identical results as they only differed by 0.0001 in the effect estimate under the scenario of medium frailty and effect heterogeneity. The two-level mixed Cox model performed worse compared to the other approaches in the presence of small heterogeneity in baseline hazards, with substantial bias in the estimated effect of 53% compared to about 1.4 to 2.9% for other approaches under this scenario, indicating over-adjustment of the mixed model when the data do not support its use. We observed similar and rather small root MSE (0.0016–0.0035) except in the case of small frailty heterogeneity and the use of a two-level mixed Cox model (~0.006 under all scenarios for pollutant effect heterogeneity). Similar results were indicated by the coverage probabilities that were generally large (>94%) except again for the case of small frailty heterogeneity and the use of a two-level mixed Cox model (87%).

4. Discussion

Our analysis supports the conclusion that effect estimates were stable under four of the evaluated approaches used to adjust for cohort (indicator variables, strata, frailty and random intercept) but substantially different when no adjustment for cohort was applied. Further

Table 3
Simulation results for the comparison of modeling approaches under the different assumptions for heterogeneity between cohorts in baseline hazard (referred to as frailty in the Table) and pollutant's effect. The mean true concentration response effect for the association between $PM_{2.5}$ and natural-cause mortality is set to 0.007.

Modeling approach		Small effect heterogeneity			Medium effect heterogeneity			Large effect heterogeneity		
		Small frailty	Medium frailty	Large frailty	Small frailty	Medium frailty	Large frailty	Small frailty	Medium frailty	Large frailty
Stratified Cox	Mean β	0.0069	0.0067	0.0066	0.0068	0.0068	0.0066	0.0069	0.0068	0.0066
	Bias	−0.0001	−0.0003	−0.0004	−0.0002	−0.0001	−0.0004	−0.0001	−0.0002	−0.0004
	Root MSE	0.0032	0.0022	0.0016	0.0032	0.0022	0.0016	0.0032	0.0022	0.0016
	Coverage probabilities	0.943	0.948	0.944	0.964	0.952	0.938	0.947	0.961	0.948
Frailty Cox	Mean β	0.0069	0.0067	0.0066	0.0068	0.0069	0.0066	0.0069	0.0068	0.0067
	Bias	−0.0001	−0.0003	−0.0004	−0.0002	−0.0001	−0.0004	−0.0001	−0.0002	−0.0003
	Root MSE	0.0032	0.0022	0.0016	0.0032	0.0022	0.0016	0.0032	0.0022	0.0016
	Coverage probabilities	0.943	0.948	0.945	0.964	0.951	0.938	0.947	0.960	0.948
Stratified Cox with random slope	Mean β	0.0069	0.0067	0.0066	0.0068	0.0068	0.0066	0.0069	0.0068	0.0067
	Bias	−0.0001	−0.0003	−0.0004	−0.0002	−0.0002	−0.0004	−0.0001	−0.0002	−0.0003
	Root MSE	0.0035	0.0024	0.0017	0.0035	0.0024	0.0017	0.0034	0.0023	0.0017
	Coverage probabilities	0.950	0.952	0.949	0.967	0.954	0.947	0.956	0.964	0.956
Mixed Cox with random intercept and slope[a]	Mean β	0.0033	0.0064	0.0066	0.0035	0.0067	0.0066	0.0037	0.0065	0.0066
	Bias	−0.0037	−0.0006	−0.0004	−0.0034	−0.0003	−0.0004	−0.0033	−0.0005	−0.0004
	Root MSE	0.0060	0.0026	0.0018	0.0057	0.0026	0.0018	0.0054	0.0025	0.0018
	Coverage probabilities	0.870	0.936	0.955	0.874	0.940	0.945	0.880	0.941	0.955

MSE: Mean Square error.
[a] With Gaussian distribution as opposed to gamma distribution for frailty Cox.

adjustment for the small area grouping in our data generally increased the effect estimates' standard errors. Simulations confirmed identical results between the stratified and frailty models, which are two methods with particular interest due to their difference in accounting for exposure contrast. Driven by these results, the main modeling approach for the pooled data analysis in ELAPSE applied a stratified multivariable Cox model to account for between-cohort heterogeneity. We did not further adjust for clustering at small area level, considering the small number of subjects at this level that also results in a small number of events. In addition, as only one covariate was available at this level, i.e. mean income, the resulted clustering was small.

Our results agree with previous methodological considerations for the choice between stratified vs frailty approach, although the stratified model does not exploit between-cohort variability in exposure and results in slightly larger CIs. Glidden & Vittinghoff (Glidden and Vittinghoff, 2004) support the use of random effects Cox models in multicenter studies, while others indicate that for large sample sizes the two approaches provide identical estimates (O'Quigley and Stare, 2002; Hosmer et al., 2008). Giganti et al. (Giganti et al., 2015), similarly to our design, reviewed seven approaches (including the ones presented here) to account for heterogeneity in Human Immunodeficiency Virus (HIV) treatment cohorts and concluded that HRs varied slightly between approaches, and differences were not clinically meaningful. Most previous publications focus on discrete exposures. Basically the two approaches have different assumptions and advantages, as the random effects models (frailty or mixed Cox) estimate between-cohort heterogeneity in underlying baseline hazards and assume a structure for these. In addition, the mixed Cox model further allows to incorporate random slopes per cohort. Therefore, heterogeneity in exposure effect can also be investigated by specifying a random effects distribution. Nevertheless, our simulations indicated stability of the estimates in the case of the small effects observed in environmental epidemiology under most heterogeneity scenarios and further argued against use of a random slope in the absence of supporting evidence as this performed worst in case of small frailty.

The choice of the modeling approach furthermore has consequences for the exposure contrast that can be exploited in the analysis. In case of fixed cohort effects and the stratified approach the analysis is largely based upon exposure contrast within cohorts, while under the random cohorts' effects, and the no cohort adjustment, we further exploit between-cohort contrasts, which could be different for different pollutants of interest. In the analyses presented here the $PM_{2.5}$ exposure contrast (on a relative scale) within cohorts was smaller than for NO_2, but possibly the large cohort sizes and number of cases counterbalance the loss of power. Basagana et al. (Basagaña et al., 2018) in their methods' comparison between fixed and random multilevel models under linear regression stressed the importance of distinguishing within- from between-studies associations as the different models exploit different contrasts. Although the stratified Cox model as the selected modeling approach in the ELAPSE project does not account for between-cohort contrast, the identical results with the frailty model (that does account for this contrast) provided by the real data application and the simulations, support its use due to its relaxing of any assumptions regarding the distribution of the underlying hazards and its easier software application. We anticipate that the similarity of the approaches is expected in the estimation of an association in pooled data from observational studies when the sample sizes are large, resulting in adequate number of cases per study. Observed differences in the hazard ratios and corresponding CIs were too small to impact the biological interpretation of the findings.

Measurement error of exposure is inherent in air pollution epidemiological studies and may bias health effect estimates. It is often considered to be non-differential, meaning that the true measurement bears no additional information on the outcome given the surrogate/proxy exposure and relevant covariates. To minimize differences in exposure assessment as a driving factor of between cohorts' heterogeneity we estimated our exposures from European-wide models instead of using cohort-specific exposure assessment methods. This does not exclude the possibility that the exposure assessment method may perform differently in different parts of the exposure distribution that could characterize different cohorts and potentially result in varying measurement error between cohorts. Nevertheless, the compared approaches account for the heterogeneity in the baseline hazards (not appropriately captured by the covariates) hence do not address measurement error impacts. Measurement error is expected to affect between cohorts exposure contrasts, but under this scenario we anticipate that the analysis not accounting for cohort would provide more similar results to the rest of the approaches. Further measurement error may lead to heterogeneous effects between cohorts, but even in this case our simulations supported the robustness of the findings and the consistency of the effects under any approach accounting for baseline hazards' heterogeneity.

5. Conclusion

In conclusion, our study shows that it is important to account for between-cohort heterogeneity in multi-center collaborations using pooled cohort data, although the specific approach may be less important. The use of a stratified approach proved optimal for the ELAPSE analyses, but we urge researchers to consider cohort-specific conditions and objectives in deciding the approach to control for different cohorts in individual participant level data in pooled analysis.

CRediT authorship contribution statement

Evangelia Samoli: Methodology, Software, Visualization, Supervision, Writing - original draft. **Sophia Rodopoulou:** Data curation, Formal analysis, Methodology, Software, Writing - review & editing. **Ulla A. Hvidtfeldt:** Formal analysis, Software, Writing - review & editing. **Kathrin Wolf:** Data curation, Formal analysis, Software, Writing - review & editing. **Massimo Stafoggia:** Methodology, Writing - review & editing. **Bert Brunekreef:** Conceptualization, Funding acquisition, Methodology, Project administration, Resources, Software, Supervision, Validation, Writing - review & editing. **Maciej Strak:** Writing - review & editing. **Jie Chen:** Data curation, Methodology, Project administration, Software, Writing - review & editing. **Zorana J. Andersen:** Data curation, Writing - review & editing. **Richard Atkinson:** Writing - review & editing. **Mariska Bauwelinck:** Data curation, Writing - review & editing. **Tom Bellander:** Writing - review & editing. **Jørgen Brandt:** Data curation, Methodology, Writing - review & editing. **Giulia Cesaroni:** Methodology, Writing - review & editing. **Francesco Forastiere:** Conceptualization, Funding acquisition, Writing - review & editing. **Daniela Fecht:** Conceptualization, Data curation, Methodology, Supervision, Writing - review & editing. **John Gulliver:** Conceptualization, Funding acquisition, Methodology, Validation, Visualization, Writing - review & editing. **Ole Hertel:** Data curation, Methodology, Writing - review & editing. **Barbara Hoffmann:** Conceptualization, Data curation, Funding acquisition, Resources, Writing - review & editing. **Kees de Hoogh:** Data curation, Methodology, Writing - review & editing. **Nicole A.H. Janssen:** Writing - review & editing. **Matthias Ketzel:** Data curation, Methodology, Writing - review & editing. **Jochem O. Klompmaker:** Writing - review & editing. **Shuo Liu:** Writing - review & editing. **Petter Ljungman:** Conceptualization, Data curation, Writing - review & editing. **Gabriele Nagel:** Data curation, Writing - review & editing. **Bente Oftedal:** Writing - review & editing. **Göran Pershagen:** Data curation, Writing - review & editing. **Annette Peters:** Data curation, Funding acquisition, Project administration, Resources, Writing - review & editing. **Ole Raaschou-Nielsen:** Conceptualization, Data curation, Methodology, Supervision, Writing - review & editing. **Matteo Renzi:** Writing - review & editing. **Doris T. Kristoffersen:** Writing - review & editing. **Gianluca Severi:** Data curation, Resources, Writing - review & editing. **Torben Sigsgaard:**

Methodology, Writing - review & editing. **Danielle Vienneau:** Writing - review & editing. **Gudrun Weinmayr:** Data curation, Methodology, Writing - review & editing. **Gerard Hoek:** Conceptualization, Funding acquisition, Methodology, Project administration, Supervision, Validation, Writing - review & editing. **Klea Katsouyanni:** Methodology, Visualization, Supervision, Writing - review & editing.

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.

Source of funding

This work was supported by Health Effects Institute (HEI) Research Agreement [grant number #4954-RFA14-3/16-5-3].

Disclaimer

"Research described in this article was conducted under contract to the Health Effects Institute (HEI), an organization jointly funded by the United States Environmental Protection Agency (EPA) (Assistance Award No. R-82811201) and certain motor vehicle and engine manufacturers. The contents of this article do not necessarily reflect the views of HEI, or its sponsors, nor do they necessarily reflect the views and policies of the EPA or motor vehicle and engine manufacturers."

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.envint.2020.106371.

References

Austin, P.C., 2017. A tutorial on multilevel survival analysis: methods, models and applications. Int. Stat. Rev. 85, 185–203.
Basagaña, X., Pedersen, M., Barrera-Gómez, J., et al., 2018. Analysis of multicentre epidemiological studies: contrasting fixed or random effects modelling and meta-analysis. Int. J. Epidemiol. 47, 1343–1354.
Beelen, R., Raaschou-Nielsen, O., Stafoggia, M., et al., 2014. 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 383, 785–795.
Bender, R., Augustin, T., Blettner, M., 2005. Generating survival times to simulate Cox proportional hazards models. Stat. Med. 24, 1713–1723.
Brunekreef, B., Strak, M., Chen, J., et al. Mortality and Morbidity Effects of Long-Term Exposure To Low-Level PM2.5, Black Carbon, NO2 and O3: An Analysis of European Cohorts. ELAPSE project: Effects of Low-Level Air Pollution: A Study in Europe. HEI Res Rep, 2021 (in press).
Cesaroni, G., Forastiere, F., Stafoggia, M., et al., 2014. Long term exposure to ambient air pollution and incidence of acute coronary events: prospective cohort study and meta-analysis in 11 European cohorts from the ESCAPE Project. BMJ 348, f7412.
de Hoogh, K., Chen, J., Gulliver, J., et al., 2018. Spatial PM2.5, NO2, O3 and BC models for Western Europe - Evaluation of spatiotemporal stability. Environ. Int. 120, 81–92.
Debray, T.P., Moons, K.G., van Valkenhoef, G., et al., 2015. Get real in individual participant data (IPD) meta-analysis: a review of the methodology. Res. Synth. Methods 6, 293–309.
Giganti, M.J., Luz, P.M., Caro-Vega, Y., et al., 2015. A Comparison of seven cox regression-based models to account for heterogeneity across multiple HIV treatment cohorts in Latin America and the Caribbean. AIDS Res. Hum. Retroviruses 31, 496–503.
Glidden, D.V., Vittinghoff, E., 2004. Modelling clustered survival data from multicentre clinical trials. Stat. Med. 23, 369–388.
Hosmer, D.H., Lemeshow, S., May, S., 2008. Applied Survival Analysis: Regression Modeling of Time-to-Event Data, second ed. John Wiley & Sons Inc, Hoboken, New Jersey.
Jolani, S., Debray, T.P.A., Koffijberg, H., van Buurenc, S., Moonsb, K.G.M., 2015. Imputation of systematically missing predictors in an individual participant data meta-analysis: a generalized approach using MICE. Stat. Med. 34, 1841–1863.
Katsouyanni, K., Touloumi, G., Samoli, E., et al., 2001. Confounding and effect modification in the short-term effects of ambient particles on total mortality: results from 29 European cities within the APHEA2 project. Epidemiology 12, 521–531.
Lindsay, C.R., Blackhall, F.H., Carmel, A., et al., 2019. EPAC-lung: pooled analysis of circulating tumour cells in advanced non-small cell lung cancer. Eur. J. Cancer 117, 60–68.
O'Quigley, J., Stare, J., 2002. Proportional hazards models with frailties and random effects. Stat. Med. 21, 3219–3233.
Pedersen, M., Giorgis-Allemand, L., Bernard, C., et al., 2013. Ambient air pollution and low birthweight: a European cohort study (ESCAPE). Lancet Respir. Med. 1, 695–704.
Raaschou-Nielsen, O., Andersen, Z.J., Beelen, R., et al., 2013. Air pollution and lung cancer incidence in 17 European cohorts: prospective analyses from the European Study of Cohorts for Air Pollution Effects (ESCAPE). Lancet Oncol. 14, 813–822.
Resche-Rigon, M., White, I.R., Bartlett, J.W., Peters, S.A.E., 2013. Thompson SG and on behalf of the PROG-IMT Study Group. Multiple imputation for handling systematically missing confounders in meta-analysis of individual participant data. Stat. Med. 32, 4890–4905.
Stafoggia, M., Cesaroni, G., Peters, A., et al., 2014. Long-term exposure to ambient air pollution and incidence of cerebrovascular events: results from 11 European cohorts within the ESCAPE project. Environ. Health Perspect. 122, 919–925.
Therneau, T., Grambsch, P., 2000. Modeling Survival Data: Extending the Cox Model. Springer, New York, NY.
Thiébaut, A.C.M., Bénichou, J., 2004. Choice of time-scale in Cox's model analysis of epidemiologic cohort data: a simulation study. Stat. Med. 23, 3803–3820.