**SUPPLEMENTARY MATERIAL**

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

Evangelia Samoli1\*, Sophia Rodopoulou1, Ulla A. Hvidtfeldt2, Kathrin Wolf3, Massimo Stafoggia4,5, Bert Brunekreef6,Maciej Strak6,7, Jie Chen6, Zorana J. Andersen8, Richard Atkinson9, Mariska Bauwelinck10, Tom Bellander5,11, Jørgen Brandt12, Giulia Cesaroni4, Francesco Forastiere13, Daniela Fecht14, John Gulliver15, Ole Hertel12, Barbara Hoffmann16, Kees de Hoogh17,18, Nicole A.H. Janssen6, Matthias Ketzel12,19, Jochem O. Klompmaker6,7, Shuo Liu8, Petter Ljungman5,20, Gabriele Nagel21, Bente Oftedal22, Göran Pershagen5,11, Annette Peters3, Ole Raaschou-Nielsen2,12, Matteo Renzi4, Doris T. Kristoffersen23, Gianluca Severi4, Torben Sigsgaard24, Danielle Vienneau17,18, Gudrun Weinmayr21, Gerard Hoek6, Klea Katsouyanni1, 13.

1Dept. of Hygiene, Epidemiology and Medical Statistics, Medical school, National and Kapodistrian University of Athens, 75 Mikras Asias Str, 115 27 Athens, Greece;

2Danish Cancer Society Research Centre, Copenhagen, Denmark.

3Institute of Epidemiology, Helmholtz Zentrum München, Neuherberg, Germany

4Department of Epidemiology, Lazio Region Health Service ASL Roma 1, Rome, Italy

5Institute of Environmental Medicine, Karolinska Institutet, Stockholm, Sweden

6Institute for Risk Assessment Sciences (IRAS), Utrecht University, Postbus 80125, 3508 TC Utrecht, the Netherlands

7National Institute for Public Health and the Environment (RIVM), Antonie van Leeuwenhoeklaan 9, 3721 MA Bilthoven, Netherlands

8University of Copenhagen, Department of Public Health, Section of Environmental Health, Øster Farimagsgade 5, 1014, Copenhagen, Denmark

9Population Health Research Institute, St George’s, University of London, Cranmer Terrace, London, SW17 0RE.

10Interface Demography, Department of Sociology, Vrije Universiteit Brussel, Brussels, Belgium.

11Centre for Occupational and Environmental Medicine, Region Stockholm, Stockholm, Sweden.

12Department of Environmental Science, Aarhus University, Frederiksborgvej 399, Roskilde, Denmark;

13NIHR HPRU Health Impact of Environmental Hazards, Environmental Research Group, Analytical, Environmental & Forensic Sciences, King's College London, UK.

14Small Area Health Statistics Unit, MRC Centre for Environment and Health, School of Public Health, Imperial College London, Norfolk Place, London, W2 1PG, UK

15Centre for Environmental Health and Sustainability & School of Geography, Geology and the Environment, University of Leicester, Leicester, UK.

16 Institute for Occupational, Social and Environmental Medicine, Medical Faculty, Heinrich-Heine-University of Düsseldorf, Germany

17Swiss Tropical and Public Health Institute, Basel, Switzerland

18University of Basel, Basel, Switzerland

19Global Centre for Clean Air Research (GCARE), University of Surrey, Guildford GU2 7XH, United Kingdom

20Department of Cardiology, Danderyd University Hospital, Stockholm, Sweden

21Institute of Epidemiology and Medical Biometry, Ulm University, Ulm, Germany

22Department of environmental health, Norwegian Institute of Public Health, Oslo, Norway

23Cluster for Health Services Research, Norwegian Institute of Public Health, Oslo, Norway

24Department of Public Health, Environment Occupation and Health, Danish Ramazzini Centre, Aarhus University, Aarhus, Denmark.

**Figure S1**. Log-log plot of the cohort-specific survival functions for natural-cause mortality.



CEANS: Cardiovascular Effects of Air Pollution and Noise in Stockholm; SDPP: The Stockholm Diabetes Preventive Program; SALT: Screening Across the Lifespan Twin Study; SNAC-K: The Swedish National Study of Aging and Care in Kungsholmen; DCH: Diet, Cancer and Health; DNC: Danish Nurses Cohort; EPIC-NL: European Prospective Investigation into Cancer and Nutrition, the Netherlands; MORGEN: Monitoring Project on Risk Factors and chronic diseases in the Netherlands; HNR: Heinz Nixdorf Recall study; 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.

**Table S1.** Description of original size and baseline characteristics of cohorts contributing data to the pooled dataset.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Cohort** | **Study Area** | **Cohort size** | **Recruitment period** | **End Year of Follow up** | **Mean Baseline age (years)** | **Female****(N(%))** | **BMI ≥ 25 kg/m2 (N(%))** | **Current smokers (N(%))** |
| CEANS-SALT | Stockholm county, Sweden | 6724 | 1998 - 2002 | 2011 | 57.8 | 3417 (55) | 2458 (40) | 1311 (21) |
| CEANS-SDPP | 7835 | 1992 - 1998 | 2011 | 47.1 | 4727 (61) | 3982 (52) | 2038 (26) |
| CEANS-SIXTY | 4180 | 1997 - 1999 | 2014 | 60.0 | 2068 (52) | 2545 (65) | 839 (21) |
| CEANS-SNACK | 3247 | 2001 - 2004 | 2011 | 72.9 | 1767 (62) | 1501 (53) | 404 (14) |
| DCH | Cities of Copenhagen and Aarhus, Denmark | 56 308 | 1993 - 1997 | 2015 | 56.7 | 28 134 (52) | 30 071 (56) | 19 459 (36) |
| DNC-1993 | Denmark-wide | 19 664 | 1993 | 2013 | 56.2 | 17 043 (100) | 4783 (28) | 6383 (37) |
| DNC-1999 | 8769 | 1999 | 2013 | 47.9 | 8128 (100) | 2439 (30) | 2325 (29) |
| E3N | France-wide | 53 507 | 1993 - 1996 | 2011 | 53.0 | 39 006 (100) | 8058 (21) | 5060 (13) |
| EPIC-NL-MORGEN | Four cities, the Netherlands | 20 683 | 1993 - 1997 | 2012 | 42.9 | 10 060 (55) | 8986 (50) | 6359 (35) |
| EPIC-NL-PROSPECT | 16 193 | 1993 - 1997 | 2013 | 57.7 | 14 570 (100) | 7978 (55) | 3335 (23) |
| HNR | Ruhr area, Germany | 4809 | 2000 - 2003 | 2015 | 59.7 | 2382 (50) | 3480 (74) | 1113 (24) |
| KORA-S3 | Augsburg area,Germany | 4556 | 1994 - 1995 | 2011 | 49.4 | 1308 (51) | 1722 (67) | 519 (20) |
| KORA-S4 | 4257 | 1999 - 2001 | 2014 | 49.3 | 1173 (51) | 1563 (69) | 523 (23) |
| VHM & PP | Vorarlberg region, Austria | 170 250 | 1985 - 2005 | 2014 | 42.1 | 81 117 (56) | 61 249 (43) | 28 916 (20) |

CEANS: Cardiovascular Effects of Air Pollution and Noise in Stockholm; SDPP: The Stockholm Diabetes Preventive Program; SALT: Screening Across the Lifespan Twin Study; SNAC-K: The Swedish National Study of Aging and Care in Kungsholmen; DCH: Diet, Cancer and Health; DNC: Danish Nurses Cohort; EPIC-NL: European Prospective Investigation into Cancer and Nutrition, the Netherlands; MORGEN: Monitoring Project on Risk Factors and chronic diseases in the Netherlands; HNR: Heinz Nixdorf Recall study; 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.

**Table S2.** Hazard ratios ((HR) and associated 95% confidence intervals (CI)) for **natural-cause mortality** associated with a 5 μg/m3 increase in PM2.5 or 10 μg/m3 increase in NO2 using several approaches for adjustment for cohort along with related Akaike Information criteria (AIC). 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 small area income. All cohorts, *n* = 325 637 subjects.

|  |  |  |
| --- | --- | --- |
| **Modeling approach** | **PM2.5** | **NO2** |
|  | **HR (95%CI)** | **AIC** | **HR (95%CI)** | **AIC** |
| No control | 1.045 (1.028, 1.062) | 962 271 | 1.035 (1.023, 1.048) | 962 270 |
| Indicator per cohort | 1.130 (1.105, 1.154) | 961 921 | 1.085 (1.069, 1.101) | 961 929 |
| Stratified by cohort | 1.130 (1.106, 1.155) | 831 515 | 1.086 (1.070, 1.102) | 831 521 |
| Frailty model  | 1.129 (1.105, 1.154) | 961 921 | 1.085 (1.069, 1.101) | 961 929 |
| Random intercept in mixed Cox | 1.125 (1.101, 1.149) | 961 918 | 1.082 (1.066, 1.098) | 961 925 |

**Table S3.** Hazard ratios ((HR) and associated 95% confidence intervals (CI)) for **incidence** of **coronary and cerebrovascular events** associated with a 5 μg/m3 increase in PM2.5 or 10 μg/m3 increase in NO2 using several approaches for adjustment for cohort along with related Akaike Information criteria (AIC). All models use age as the time axis, stratify by sex and control for body mass index, employment and marital status, education, smoking status, smoking intensity and duration and mean small area income. All cohorts except VHM&PP and E3N, *n* = 137 148 subjects.

|  |  |  |
| --- | --- | --- |
| **Modeling approach** | **PM2.5** | **NO2** |
| **Coronary events** | **HR (95%CI)** | **AIC** | **HR (95%CI)** | **AIC** |
| No control | 1.014 (0.976, 1.054) | 202 232 | 1.004 (0.979, 1.029) | 202 232 |
| Indicator per cohort | 1.021 (0.945, 1.103) | 201 738 | 1.039 (1.008, 1.071) | 201 732 |
| Stratified by cohort | 1.021 (0.945, 1.103) | 174 176 | 1.039 (1.008, 1.071) | 174 170 |
| Frailty model  | 1.020 (0.945, 1.101) | 201 738 | 1.039 (1.008, 1.071) | 201 732 |
| Random intercept in mixed Cox | 1.016 (0.943, 1.094) | 201 738 | 1.039 (1.009, 1.071) | 201 732 |
| **Cerebrovascular events** | **HR (95%CI)** | **AIC** | **HR (95%CI)** | **AIC** |
| No control | 0.847 (0.809, 0.886) | 138 715 | 0.994 (0.965, 1.025) | 138 767 |
| Indicator per cohort | 1.100 (1.005, 1.205) | 138 440 | 1.075 (1.036, 1.115) | 138 430 |
| Stratified by cohort | 1.103 (1.008, 1.208) | 121 695 | 1.076 (1.037, 1.116) | 121 684 |
| Frailty model  | 1.095 (1.001, 1.198) | 138 440 | 1.075 (1.036, 1.115) | 138 430 |
| Random intercept in mixed Cox | 1.069 (0.981, 1.164) | 138 440 | 1.072 (1.034, 1.112) | 138 429 |

E3N: Etude Epidémiologique auprès de femmes de la Mutuelle Générale de l'Education Nationale; VHM&PP: Vorarlberg Health Monitoring and Prevention Programme.

**Table S4.** Hazard ratios ((HR) and associated 95% confidence intervals (CI)) for **lung cancer incidence** associated with a 5 μg/m3 increase in PM2.5 or 10 μg/m3 increase in NO2 using several approaches for adjustment for cohort along with related Akaike Information criteria (AIC). 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 small area income. All cohorts, except KORA. *n* = 307 550 subjects.

|  |  |  |
| --- | --- | --- |
| **Modeling approach** | **PM2.5** | **NO2** |
|  | **HR (95%CI)** | **AIC** | **HR (95%CI)** | **AIC** |
| No control | 1.167 (1.101, 1.236) | 80 801 | 0.978 (0.939, 1,019) | 80 827 |
| Indicator per cohort | 1.140 (1.053, 1.235) | 80 636 | 1.030 (0.980, 1.082) | 80 645 |
| Stratified by cohort | 1.132 (1.045, 1.226) | 69 478 | 1.020 (0.970, 1.072) | 69 487 |
| Frailty model  | 1.139 (1.053, 1.233) | 80 635 | 1.029 (0.979, 1.081) | 80 645 |
| Random intercept in mixed Cox | 1.132 (1.050, 1.221) | 80 634 | 1.026 (0.977, 1.077) | 80 643 |

KORA: Kooperative Gesundheits forschung in der Region Augsburg.

**Table S5.** Hazard ratios ((HR) and associated 95% confidence intervals (CI)) for **incidence of coronary and cerebrovascular events** associated with a 5 μg/m3 increase in PM2.5 or 10 μg/m3 increase in NO2 using several approaches for adjustment for cohort and small area (SA) level. All models use age as the time axis, stratify by sex and control for body mass index, employment and marital status, education, smoking status, smoking intensity and duration and mean SA income. All cohorts except DNC, VHM&PP and E3N, *n* = 112 283 subjectsa.

|  |  |  |
| --- | --- | --- |
|  | **Coronary Events** | **Cerebrovascular Events** |
| **Modeling approach**  | **PM2.5** **HR (95%CI)** | **NO2** **HR (95%CI)** | **PM2.5** **HR (95%CI)** | **NO2 effect****HR (95%CI)** |
| No control for cohort |
|  No SA control  | 1.062 (1.019, 1.106) | 1.082 (1.05, 1.114) | 0.838 (0.799, 0.879) | 0.987 (0.953, 1.022) |
|  Variance correction  | 1.062 (0.988, 1.141) | 1.082 (1.035, 1.131) | 0.838 (0.786, 0.893) | 0.987 (0.923, 1.056) |
| Indicator for cohort |
|  No SA control  | 1.067 (0.974, 1.17) | 1.058 (1.021, 1.097) | 1.092 (0.986, 1.209) | 1.079 (1.035, 1.125) |
|  Variance correction  | 1.067 (0.956, 1.192) | 1.058 (1.023, 1.094) | 1.092 (0.967, 1.233) | 1.079 (1.041, 1.119) |
| Strata for cohort |
|  No SA control  | 1.065 (0.972, 1.168) | 1.057 (1.02, 1.096) | 1.093 (0.986, 1.21) | 1.079 (1.035, 1.126) |
|  Variance correction | 1.065 (0.956, 1.187) | 1.057 (1.022, 1.094) | 1.093 (0.969, 1.231) | 1.079 (1.042, 1.118) |
| Mixed Cox |
|  No SA control  | 1.05 (0.966, 1.141) | 1.061 (1.024, 1.099) | 1.061 (0.963, 1.168) | 1.077 (1.034, 1.123) |
|  Two levels adjustment for cohort and SA | 1.033 (0.939, 1.136) | 1.05 (1.006, 1.096) | 1.05 (0.951, 1.159) | 1.077 (1.033, 1.123) |

a Models with smaller sample size due to missing values for small area characterization; *n* in Table S3 = 137 148; *n* in Table S5 = 112 283.

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.

**Table S6.** Hazard ratios ((HR) and associated 95% confidence intervals (CI)) for **lung cancer incidence** associated with a 5 μg/m3 increase in PM2.5 or 10 μg/m3 increase in NO2 using several approaches for adjustment for cohort and small area (SA) level. 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, E3N and KORA. *n* = 107 663 subjectsa.

|  |  |  |
| --- | --- | --- |
| **Modeling approach** | **PM2.5** **HR (95%CI)** | **NO2** **HR (95%CI)** |
| No control for cohort |
|  No SA control  | 1.033 (0.943, 1.132) | 0.984 (0.928, 1.043) |
|  Variance correction  | 1.033 (0.946, 1.128) | 0.984 (0.933, 1.038) |
| Indicator for cohort |
|  No SA control  | 1.103 (0.935, 1.303) | 0.999 (0.932, 1.072) |
|  Variance correction  | 1.103 (0.998, 1.219) | 0.999 (0.943, 1.059) |
| Strata for cohort |
|  No SA control  | 1.089 (0.922, 1.287) | 0.993 (0.926, 1.065) |
|  Variance correction | 1.089 (0.989, 1.200) | 0.993 (0.935, 1.055) |
| Mixed Cox |
|  No SA control  | 1.091 (0.955, 1.246) | 1.008 (0.943, 1.077) |
|  Two levels adjustment for cohort and SA | 1.092 (0.955, 1.248) | 1.008 (0.943, 1.077) |

a Models with smaller sample size due to missing values for small area characterization; *n* in Table S4 = 307 550; *n* in Table S6 = 107 663.

DNC: Danish Nurses Cohort; 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.

**Table S7.** Hazard ratios ((HR) and associated 95% confidence intervals (CI)) for **natural-cause mortality** associated with a 5 μg/m3 increase in PM2.5 or 10 μg/m3 increase in NO2 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. All cohorts except DNC, VHM&PP and E3N, *n* = 116 807 subjectsa.

|  |  |  |
| --- | --- | --- |
| **Modeling approach** | **PM2.5****HR (95%CI)** | **NO2****HR (95%CI)** |
| No control for cohort |
|  No control for SA | 1.013 (0.987, 1.039) | 1.038 (1.018, 1.058) |
|  Variance correction for SA | 1.013 (0.974, 1.053) | 1.038 (0.992, 1.085) |
| Indicator for cohort |
|  No control for SA | 1.197 (1.127, 1.271) | 1.067 (1.041, 1.093) |
|  Variance correction for SA | 1.197 (1.086, 1.320) | 1.067 (1.026, 1.109) |
| Strata for cohort |
|  No control for SA | 1.187 (1.118, 1.261) | 1.061 (1.036, 1.088) |
|  Variance correction for SA | 1.187 (1.072, 1.316) | 1.061 (1.020, 1.105) |
| Mixed Cox |
|  No control for SA | 1.181 (1.115, 1.251) | 1.068 (1.042, 1.093) |
|  Two levels adjustment for cohort and SA | 1.064 (1.000, 1.133) | 1.033 (1.002, 1.065) |

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.

**Extracts from code used in R (version 3.4.0) to apply the approaches**

**for the association between natural-cause mortality and NO2**

**#### Libraries ####**

library(survival) # version 2.42-3

library(coxme) # version 2.2-10

**#### Import the data ####**

load("dataset.rdata")

1. ***Approaches to account for first level of clustering in our pooled dataset corresponding to cohort specification***

*#### Cox model without control for cohort indicated by variable named “studyid” ####*

model0.no2<-coxph(Surv(age\_b,age\_end,nat\_mort)~year+strata(sex)+as.factor(smoking)+

smokeint+smokeint2+smokedur+as.factor(bmi\_cat)+as.factor(mar\_stat)+as.factor(employed)+meaninc\_neighbor\_2001+NO2,data=dataset)

*#### Cox model with dummies for cohort (“studyid”) ####*

model1.no2<-coxph(Surv(age\_b,age\_end,nat\_mort)~year+strata(sex)+as.factor(smoking)+

smokeint+smokeint2+smokedur+as.factor(bmi\_cat)+as.factor(mar\_stat)+as.factor(employed)+ meaninc\_neighbor\_2001+as.factor(studyid)+NO2,data=dataset)

*#### Cox model also stratified by cohort (“studyid”) ####*

model2.no2<-coxph(Surv(age\_b,age\_end,nat\_mort)~year+strata(studyid,sex)+

as.factor(smoking)+smokeint+smokeint2+smokedur+as.factor(bmi\_cat)+as.factor(mar\_stat)+ as.factor(employed)+meaninc\_neighbor\_2001+NO2,data=dataset)

*#### Cox model with shared frailty for cohort (“studyid”) ####*

model3.no2<-coxph(Surv(age\_b,age\_end,nat\_mort)~year+strata(sex)+as.factor(smoking)+

smokeint+smokeint2+smokedur+ as.factor(bmi\_cat)+as.factor(mar\_stat)+as.factor(employed)+

meaninc\_neighbor\_2001+frailty(studyid,sparse=FALSE)+NO2,data=dataset)

*### Mixed Cox model with random intercept for cohort (“studyid”) ####*

model4.no2<-coxme(Surv(age\_b,age\_end,nat\_mort)~year+strata(sex)+as.factor(smoking)+

smokeint+smokeint2+smokedur+as.factor(bmi\_cat)+as.factor(mar\_stat)+as.factor(employed)+ meaninc\_neighbor\_2001+NO2+(1|studyid),data=dataset)

1. ***Approaches to account for the two levels of clustering in our pooled dataset: one corresponding to cohort specification and the second to the small area level***

*#### Cox model without control for cohort (“studyid”) plus cluster term for small area indicated by variable named “neighb\_id” ####*

model0.1.no2<-coxph(Surv(age\_b,age\_end,nat\_mort)~year+strata(sex)+as.factor(smoking)+

smokeint+smokeint2+smokedur+as.factor(bmi\_cat)+as.factor(mar\_stat)+

as.factor(employed)+meaninc\_neighbor\_2001+cluster(neighb\_id)+NO2,data=dataset)

*#### Cox model with dummies for cohort (“studyid”) plus cluster term for small area (“neighb\_id”) ####*

model1.1.no2<-coxph(Surv(age\_b,age\_end,nat\_mort)~year+strata(sex)+as.factor(smoking)+

smokeint+smokeint2+smokedur+as.factor(bmi\_cat)+as.factor(mar\_stat)+as.factor(employed)+meaninc\_neighbor\_2001+as.factor(studyid)+cluster(neighb\_id)+NO2,data=dataset)

*#### Cox model also stratified by cohort (“studyid”) plus cluster term for small area (“neighb\_id”) ####*

model2.1.no2<-coxph(Surv(age\_b,age\_end,nat\_mort)~year+strata(studyid,sex)+as.factor(smoking)+

smokeint+smokeint2+smokedur+as.factor(bmi\_cat)+as.factor(mar\_stat)+

as.factor(employed)+meaninc\_neighbor\_2001+cluster(neighb\_id)+ NO2,data=dataset)

*#### Mixed Cox model with random intercept for two levels of clustering: 1) for cohort (“studyid”) and 2) for small area nested within cohort (“neighb\_id”) ####*

model4.1.no2<-coxme(Surv(age\_b,age\_end,nat\_mort)~year+strata(sex)+as.factor(smoking)+

smokeint+smokeint2+smokedur+as.factor(bmi\_cat)+as.factor(mar\_stat)+as.factor(employed)+meaninc\_neighbor\_2001+NO2+(1|studyid/neighb\_id),data=dataset)