Environment International 126 (2019) 762–770

Contents lists available at ScienceDirect

Environment International

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

All-cause mortality and long-term exposure to low level air pollution in the '45 and up study' cohort, Sydney, Australia, 2006–2015

Ivan C. Hanigan[a,b,c,*], Margaret I. Rolfe[b], Luke D. Knibbs[a,d], Farhad Salimi[a,b], Christine T. Cowie[a,e,f], Jane Heyworth[a,g], Guy B. Marks[a,h], Yuming Guo[a,i], Martin Cope[a,j], Adrian Bauman[k], Bin Jalaludin[a,l,m], Geoffrey G. Morgan[a,b]

[a] Centre for Air Pollution, Energy and Health Research, Australia
[b] The University of Sydney, University Centre for Rural Health, School of Public Health, Sydney, Australia
[c] Centre for Research and Action in Public Health, University of Canberra, Canberra, Australia
[d] School of Public Health, The University of Queensland, Herston, Australia
[e] South West Sydney Clinical School, University of NSW, Australia
[f] Ingham Institute for Applied Medical Research, Woolcock Institute of Medical Research, The University of Sydney, Sydney, Australia
[g] The Clean Air and Urban Landscapes Hub & School of Population and Global Health, The University of Western Australia, Australia
[h] Woolcock Institute of Medical Research & South West Sydney Clinical School, University of New South Wales, Sydney, Australia
[i] Department of Epidemiology and Preventive Medicine, School of Public Health and Preventive Medicine, Monash University, Melbourne, Australia
[j] CSIRO, Melbourne, Australia
[k] The University of Sydney, School of Public Health, Sydney, Australia
[l] Woolcock Institute of Medical Research, University of Sydney, Sydney, Australia
[m] School of Public Health and Community Medicine, University of New South Wales & Ingham Institute for Applied Medical Research, Sydney, Australia

ARTICLE INFO

Handling Editor: Hanna Boogaard

Keywords:
All-cause mortality
Low concentration
Fine particulate matter
Gaseous pollutants
Survival model

ABSTRACT

Background: Epidemiological studies show that long-term exposure to ambient air pollution reduces life expectancy. Most studies have been in environments with relatively high concentrations such as North America, Europe and Asia. Associations at the lower end of the concentration-response function are not well defined.

Objectives: We assessed associations between all-cause mortality and exposure to annual average particulate matter < 2.5 μm ($PM_{2.5}$) and nitrogen dioxide (NO_2) in Sydney, Australia, where concentrations are relatively low.

Methods: The '45 and Up Study' comprises a prospective longitudinal cohort from the state of New South Wales, Australia with 266,969 participants linked to death registry data. We analyzed data for the participants who resided in Sydney at baseline questionnaire (n = 75,268). Exposures to long-term pollution were estimated using annual averages from a chemical transport model ($PM_{2.5}$), and a satellite-based land-use regression model (NO_2). Socio-demographic information was extracted from the baseline questionnaire. Cox proportional hazard models were applied to estimate associations, while adjusting for covariates.

Results: In our cohort mean annual $PM_{2.5}$ was 4.5 μg/m^3 and mean NO_2 was 17.8 μg/m^3. The mortality rate was 4.4% over the 7 years of follow up. Models that adjusted for individual-level and area-level risk factors resulted in a detrimental non statistically significant hazard ratio (HR) of 1.05 (95% CI: 0.98–1.12) per 1 μg/m^3 increase in $PM_{2.5}$, and 1.03 (95% CI: 0.98–1.07) per 5 μg/m^3 increase in NO_2.

Conclusions: We found evidence that low-level air pollution exposure was associated with increased risk of mortality in this cohort of adults aged 45 years and over, even at the relatively low concentrations seen in Sydney. However, a clear determination of the association with mortality is difficult because the results were sensitive to some covariates. Our findings are supportive of emerging evidence that exposure to low levels of air pollution reduces life expectancy.

* Corresponding author at: Centre for Air Pollution, Energy and Health Research, Australia.
E-mail address: ivan.hanigan@sydney.edu.au (I.C. Hanigan).

https://doi.org/10.1016/j.envint.2019.02.044
Received 3 December 2018; Received in revised form 5 February 2019; Accepted 16 February 2019
Available online 15 March 2019

1. Introduction

Associations between long-term exposure to ambient air pollution and premature mortality have been consistently shown since the early work of the Harvard Six Cities Study (Dockery et al., 1993) and the American Cancer Society Study (Pope III et al., 1995). A meta-analysis of the European Study of Cohorts for Air Pollution Effects (ESCAPE) (Beelen et al., 2014) showed strong associations with increased all-cause mortality and air pollutants especially particulate matter < 2.5 μm ($PM_{2.5}$) (also see Chen et al., 2008) and nitrogen dioxide (NO_2) (also see Hoek et al., 2013). While pooled analyses have been used to quantify the exposure-response function across the concentration spectrum (Burnett et al., 2018; Burnett et al., 2014), evidence is lacking at the lower end of this function because most existing cohorts have been from locations where pollutant levels are relatively high. Recently more cohort studies with relatively low $PM_{2.5}$ exposures have been published, for example one such study showed that there was a steeper slope on the regression line found in a 'low exposure' subgroup of the 61 million American Medicare population cohort (with exposures under 12 μg/m^3) than that observed in a model of the whole cohort (Di et al., 2017). Another study from Canada (mean = 6.3 μg/m^3) observed an increased risk of all non-accidental mortality and did not find evidence for any threshold of a 'safe' lower level of pollution (Pinault et al., 2016). Integrated exposure-response functions (IER) have been developed to enable quantification of the health effect of outdoor air pollution over the full range of exposures and currently have an assumed theoretical minimum risk level (Burnett et al., 2014). Studies from low exposure environments are needed to support the refinement of the IER at the low end.

The multicenter ESCAPE group describe 22 cohorts in Europe where observed $PM_{2.5}$ levels across cohorts ranged from 7.7 to 31.0 μg/m^3 (12 cohorts were > 10 μg/m^3), while NO_2 ranged from 5.2 to 59.8 μg/m^3 with the majority (13 cohorts) having levels > 20 μg/m^3 (Beelen et al., 2014). In comparison the regulatory (background) monitoring sites in Sydney had lower annual averages of 5.5 μg/m^3 for $PM_{2.5}$ and 14.3 μg/m^3 for NO_2 observed in 2011 (NSW Office of Environment and Heritage, n.d.). Annual average NO_2 concentrations in 15 cohort studies (from North America, Europe, China and Japan) included in a 2013 review (Hoek et al., 2013) ranged from 17.0 to 67.0 μg/m^3, with a mean 35.8 μg/m^3. The cohort in Rome, a city with a similar population size to Sydney, had an annual average NO_2 of 44.0 μg/m^3, three times higher than the average for NO_2 concentration measured in Sydney in 2011.

The IER functions are used to estimate the global burden of disease (GBD) attributable to air pollution. The GBD study estimated that in 2015 exposure to ambient $PM_{2.5}$ air pollution was the sixth-leading risk factor for burden of disease globally (Forouzanfar et al., 2016). The IER functions developed by Burnett et al. (2014) were extended and used by Apte et al. (2015) and Cohen et al. (2017) to assess how regional and global improvements in ambient air quality could reduce attributable mortality. The IERs used by Apte et al. (2015) have a low exposure "threshold" for $PM_{2.5}$ distributed between 5.8 and 8.0 μg/m^3. This lower bound for the theoretical minimum risk level is higher (and thus more conservative) than those used in recent GBD work (Forouzanfar et al., 2016) that updated information for the GBD 2015 estimates and defined a $PM_{2.5}$ threshold distribution between 2.4 and 5.9 μg/m^3. Conservative decisions regarding the level of theoretical minimum risk are usually taken because of a lack of empirical studies to support estimates of risk at low concentrations. However, the absence of such evidence does not imply evidence of the absence of risk at low concentrations, and so further research is required in low concentration environments such as Sydney.

In this study, we aimed to contribute to the evidence base on long-term exposure to low air pollution concentrations. We analyzed a large cohort of adults sampled from the general population aged 45 years and over from the Australian '45 and Up Study' cohort, the largest general population cohort in the southern hemisphere (Mealing et al., 2010). Improved understanding of the health risks at lower levels of air pollution exposure will inform current and future air pollution standards, as well as monitoring, policy and management programs to protect population health.

2. Methods

2.1. Study population

We drew our study population from the '45 and Up Study', a prospective longitudinal cohort from the state of New South Wales (NSW), Australia, within which Sydney is located. Each participant's baseline address was geocoded to longitude and latitude to enable linkage with neighborhood socio-demographic factors and environmental exposures.

Participants aged 45 years or older were randomly sampled from the Department of Human Service enrolment database, which is Australia's universal health care system and provides near complete coverage of the population. All NSW residents over 45 years at time of recruitment were eligible to be sampled. The overall response rate to the baseline questionnaire was estimated to be approximately 18% (Banks et al., 2008). The response rate is variable across the different age/gender groups sampled because people aged 80 years and older were oversampled by a factor of two, and males aged 80 years or over were oversampled compared to females (Mealing et al., 2010). Participants completed the baseline questionnaire (between 2006 and 2009) and gave consent for follow-up via linkage of their personal information to the death registry. We included participants if their address was inside the Sydney metropolitan area as this is the extent of our air pollution model domain and we could ensure comparability of air pollution exposures within the population subset being studied. We also restricted the age of participants included in our study population to those who were younger than 80 years of age at recruitment for two reasons, i) the oversampling of participants aged 80 years or older resulting in a bimodal age distribution and may lead to selection bias due to the "healthy volunteer effect" (Lindsted et al., 1996) as a greater proportion of participants would be aged older than the average life expectancy of the general population (Beadle et al., 2013) and ii) for comparability with international studies, which generally use age groups < 80 years (Beelen et al., 2014). Therefore, the sample population base for our study was any cohort participants within the Sydney metropolitan airshed who were aged 45 to 79 at recruitment.

The data collection of the '45 and Up Study' was approved by the University of New South Wales Human Research Ethics Committee (HREC). The study presented in this paper was approved by the NSW Population & Health Services Research Ethics Committee (reference: HREC/15/CIPHS/4), the Cancer Institute NSW (reference: 2015/02/575 Air Pollution, Traffic Exposures and Mortality and Morbidity in Older Australians (APTEMA) Study).

2.2. Mortality data

Mortality data from 2007 to 2015 was extracted from the NSW Register of Births Deaths and Marriages (RBDM) and linked to '45 and Up Study' participants. This data included date and age of death and other individual level information but not cause of death coding. The Centre for Health Record Linkage (CHeReL) linked the cohort and death registration data for this study.

2.3. Air pollution exposure

We estimated air pollutant concentrations in the neighborhood of the coordinates of the baseline addresses of the study participants. For $PM_{2.5}$ we used a chemical transport model (CTM) with estimated air pollutant concentration fields from a meteorological and air emissions model. A CTM uses numerical simulation algorithms to produce the

estimated concentrations. An existing $PM_{2.5}$ CTM modelled surface for the year between July 2010 to June 2011 was used (Cope et al., 2014). We assessed the temporal trends in $PM_{2.5}$ and consider the spatial pattern in this year to be representative of the annual averages in previous years. The CTM estimates were computed for a grid of 1 km × 1 km cells over a 100 km × 100 km region covering the Sydney region. This dataset was then "blended" to incorporate information from fixed site monitor data and adjusted the modelled estimates to match the observations (Physick et al., 2007). The cross-validated leave one out average R^2 for a 24 h time-step model run over the year July 2010–June 2011 was 0.70 and the model captures many of the observed daily $PM_{2.5}$ peaks. An additional model run used in sensitivity analyses was completed for a larger 300 km × 300 km region with grid cells at 3 km × 3 km resolution but these were not blended due to lack of available monitor data.

For NO_2, we used estimated concentrations for 2007 from a spatial regression model using satellite and land use data (Knibbs et al., 2014; Knibbs et al., 2016). Knibbs et al. (2014) found that year to year differences were small for the NO_2 model between 2006 and 2011 so we assumed the 2007 data were representative of long-term exposures. In an independent validation at 98 measurement sites the R^2 was 0.69 using annual ambient NO_2 concentrations 2006–2011 (Knibbs et al., 2016). We used this model to estimate annual average NO_2 concentration at Australian Bureau of Statistics (ABS) mesh blocks. Mesh blocks are the smallest geographical area defined by ABS and contains around 30–60 dwellings. Participants' exposure to annual average NO_2 in 2007 was calculated at their mesh block centroid.

2.4. Individual-level demographic and risk factor data

Information on the participants was collected in the baseline questionnaire and included demographic data on age, education, ethnicity, marital status; lifestyle and habits including physical activity, smoking and alcohol consumption; current medications; history of disease; surgical procedures; psychological distress; social support and employment status, paid and unpaid work and income. Detailed definitions of all the variable names of the '45 and Up Study' baseline data are available from the website: https://www.saxinstitute.org.au/our-work/45-up-study/data-book/. The baseline survey items are well suited to support our aim to understand the impact of air pollution while adjusting for relevant social, economic and behavioral factors on the health of Australians in mid to later life.

2.5. Area level socio-economic status (SES)

Neighborhood-level socio-economic status (SES) data was attached to participant records, using the ABS's 2006 Socio-Economic Indexes for Areas Index of Relative Socio-Economic Disadvantage (SEIFA IRSD). This index was based on 16 different measures such as the proportion of residents who had low income, low education, were unemployed, were separated/divorced, single parents and unskilled occupations derived from 2006 Census variables at a Collector District (CD) spatial area. CDs are the smallest available area in the Australian Standard Geographic Classification (ASGC) and the smallest area for which the 2006 SEIFA indexes are available. The IRSD score has a mean of 1000 and standard deviation of 100 across the entire country, but we re-expressed these as quartiles within the Sydney sample. Low scores (quartile 1) below 1000 indicate the most disadvantage relative to all cohort participants, and higher scores indicate less relative disadvantage (quartile one < 1000, quartile two = 1000–1072, quartile three = 1073–1119, and quartile four ≥1120).

2.6. Statistical analyses

We first assessed correlation between variables in exploratory data analysis using Pearson's *r* for interval data, Spearman's ρ for ordinal or ANOVA for nominal variables. We then used survival analysis (Cox proportional hazards models) to quantify associations between the two pollutants $PM_{2.5}$ and NO_2 and all-cause mortality in single-pollutant models, adjusting for other risk factor covariates and potential confounders. We used age as the time variable as this is expected to provide better adjustment for potential time-dependent confounding in the model (Shariff et al., 2008). The survival models also included a left-truncation adjustment for age at recruitment (Hanley and Foster, 2014). Subjects were censored at the day of death, while subjects who had not died by 31 March 2015 were censored on that day (the last day when linked death registry information was available).

2.7. Candidate models

Potential effect modifiers and confounders were identified from published literature, particularly the pooled analysis of ESCAPE cohorts reported by Beelen et al. (2014). All individual-level covariates were included as categories, which assisted in modelling non-linear relationships of the covariates with mortality. We adjusted for variables chosen a priori based on their potential to be confounders or effect modifiers following the ESCAPE cohorts study protocol (ESCAPE Statistical Tasks Working Group, 2012) because we wanted to ensure direct comparison with the other studies (Beelen et al., 2014; Cesaroni et al., 2014; Raaschou-Nielsen et al., 2013). Model A included age as the time axis, sex, and calendar time of enrolment (2006–2007 or 2008–2009). Model B added individual level variables excluding variables that could be on the pathway linking air pollution to health outcome. The variables in Model B included smoking status (never, former, current), weekly alcohol consumption (none, 1–7, 8–14, > 14 standard drinks), body-mass index (BMI, as underweight, normal, overweight, or obese), educational level (low = no school certificate, school certificate, or intermediate certificate; medium = high school certificate, trade/apprenticeship or diploma; high = university degree or higher), working status (employed = working full-time, part-time or self-employed; not employed = retired, home duties, unemployed, student, unpaid work), and whether they were born in Australia (yes/no).

Model C added marital status and physical activity to Model B. Marital status was considered a distal causal influence on the more proximal health risk factors such as smoking, alcohol and BMI and therefore potentially an important risk factor. This variable was withheld from Model B because of concerns about overadjustment. We observed a correlation between marital status and exposure in our study participants from exploratory data analysis (shown in Supplemental Material Table S1) so we included this in Model C to compare our results without this correlated variable. Marital status was classified as individuals who were single/divorced/widowed/separated compared with those who were married/de facto. We withheld physical activity from Model B because this is potentially a confounder or mediator as people who live in polluted areas may not exercise as much, especially if they suffer from air pollution related diseases that reduce their physical abilities, and physical inactivity is associated with mortality. We therefore also included sufficient physical activity in Model C (considered sufficient if activity occurs 5 or more times per week and total duration ≥ 150 min).

Model D (considered to be our main model) extended Model C with the addition of area-level SES and a random intercept for the different statistical local areas (SLA) from the 2006 census. The random intercept is to accommodate any residual spatial clustering. Only participants with complete data for all models A–D were used in the primary analyses.

Model diagnostics were checked for Model B and modifications were made to subsequent models accordingly. For example, stratification by covariates was used to resolve issues of their non-proportional hazards and the influence of any high leverage points (e.g. with extremely high exposure estimates) was assessed by re-fitting models with

exposures truncated to the limits of data density. Tests for statistical significance were set at $P \leq 0.05$. The R environment for statistics and graphics version 3.4.1 was used for all analyses (R Core Team, 2017).

2.8. Main model and sensitivity analyses

We considered Model D as our main model based on the inclusion of individual and area-level risk factors known from prior knowledge, and because of consistency with the ESCAPE study protocol used by many other cohort studies. In sensitivity analyses, we compared our result for Model D when missing predictor data were imputed using a recursive partitioning model. We then explored the possibility of a difference in mortality by gender (because of possible effect modification due to biological or behavioral risk factors). We also assessed the potential for non-linear exposure-response curves. Possible non-linearity in the exposure-response function was assessed using splines added with three, four and five degrees of freedom and by assessing the difference in the Bayesian Information Criterion (BIC). We chose to use BIC because it is considered more appropriate than the Akaike Information Criterion (AIC) when the aim of modelling is explanation (Shmueli, 2010). AIC is the preferred test when modelling aims to build a predictive model but can lead to overfitting and was not used for this reason. We used a threshold of six-point difference between BIC to define a strong evidence for an improvement (Raftery, 1995). We also fit additional models to an expanded study region for which we had exposures (albeit with lower confidence in the accuracy of the estimates) based on CTM model domain at 3 km × 3 km gridded resolution. The difference in BIC was used to assess the support for model re-parametrisations such as the addition of interaction terms or non-linear splines (Shmueli, 2010). Finally, we assessed the inclusion of ozone in the NO_2 model as this is a potential confounder because NO_2 is a precursor to ozone formation. Further information is presented in the Supplemental Material.

3. Results

Of the 85,846 cohort participants who met our study inclusion criteria for location within the Sydney metropolitan airshed and age between 45 and 79 years old at recruitment we had complete covariate data across all models for 75,268 participants for the main $PM_{2.5}$ survival models (i.e. about 12% missing covariate data for Model D). From this study population 3282 participants had died during the 7-year (average) period of follow up (4.4%). After excluding those with missing NO_2 estimates from this dataset there were 75,145 people with 3280 deaths.

Fig. 1 shows a map of the study area with air pollution exposure estimates. Panel A includes $PM_{2.5}$ annual averages (for our inner study region with 1 km × 1 km pixels) and panel B includes NO_2 annual averages for the same region. The concentrations are generally highest in the central parts of the city, around major roads, and near industrial facilities.

Table 1 summarizes the study subject characteristics at baseline and Table 2 shows the distribution of the cohort pollutant exposures.

$PM_{2.5}$ and NO_2 were highly correlated (0.73, $P < 0.001$) and therefore were not included in the same model. Correlations with air pollutants and all covariates are shown in the Supplemental Material. Pearson correlations for interval data are shown in Fig. S1, Spearman correlations for ordinal data are in Fig. S2 and ANOVA eta for nominal data are in Table S1. Both pollutants were negatively correlated with increasing area-level socio-economic status quartile (Spearman $PM_{2.5}$ -0.16 $P < 0.001$, NO_2 -0.12 $P < 0.001$) indicating higher pollution with higher disadvantage. There were correlations between both pollutants and marital status (ANOVA eta for $PM_{2.5} = 0.11$, $P < 0.001$ and for $NO_2 = 0.14$, $P < 0.001$) with single people having higher average exposure than partnered people. There were also correlations between both pollutants and being born in Australia (ANOVA eta for both $PM_{2.5}$ and $NO_2 = 0.09$, $P < 0.001$) with people born in Australia having lower average exposure than those born overseas.

Cohort average $PM_{2.5}$ and NO_2 exposure were generally lower than comparable cohorts from around the world. In our cohort $PM_{2.5}$ annual average was around 4.5 μg/m^3 and NO_2 average of 17.8 μg/m^3. Fig. 2 shows histograms of the cohort exposures.

Table 3 shows the results of our statistical models of all-cause mortality and exposure to air pollution for models A, B, C and D. Hazard ratios (HRs) with 95% confidence intervals (CIs) are expressed for 1 μg/m^3 change in $PM_{2.5}$ and 5 μg/m^3 change in NO_2.

All models show that both pollutants had an adverse effect on mortality (i.e., HR estimates all greater than one), as shown in Table 3. Models A and B had statistically significant associations between both pollutants and all-cause mortality. Associations in Models C and D were above one but non statistically significant. We considered Model D as our main model because in our assessment it should adequately adjust for individual and area-level risk factors based on previous literature. Full results for all covariates included in Model D are shown in Supplemental Material Table S2. The Model D HR was 1.05 (95% CI 0.98–1.12) for each 1 μg/m^3 increase in $PM_{2.5}$ exposure and 1.03 (95% CI 0.98–1.07) for each 5 μg/m^3 increase in NO_2 (in separate single pollutant models). Including marital status and sufficient physical activity (Model C) and area-level SES (Model D) reduced the magnitude and statistical significance of HR for both pollutants compared to Model B.

We conducted several sensitivity analyses and present the additional details in the Supplemental Material. Briefly, we assessed results for Model D when missing data were predicted using a recursive partitioning method. This increased the number of participants to the full sample of 85,846. The HR for $PM_{2.5}$ was reduced to 1.02 (95%CI 0.96–1.09) while the HR for NO_2 was slightly reduced (1.02, 95%CI 0.99–1.06). The interaction between gender and pollutants did not reach significance (assessed using the difference in BIC). However, the difference in the coefficients for each gender subgroup indicated potential differences, with an estimated HR for males of 1.08 (95% CI 1.00–1.17) per 1 μg/m^3 increase in $PM_{2.5}$ and an HR of 1.05 (95% CI 1.00–1.10) for a 5 μg/m^3 increase in NO_2. No strong HR was observed among females in this interaction model. The addition of a location random effect increased the HRs slightly, compared to their equivalents without random effects however the random effect term was non-significant in Model D for both pollutants (p-value ~0.1 for both). When we assessed non-linearity in the exposure-response functions, neither pollutant exhibited a significant non-linear exposure-mortality relation (tested using splines with 3,4 or 5 knots assessed using difference in BIC). The inclusion of ozone in the NO_2 model to account for potential confounding by this pollutant did not find any notable change in the point estimate or confidence intervals for NO_2. Finally, extending the study area to a 300 km × 300 km area around Sydney nearly doubled the number of study subjects compared with the primary analysis. The HRs for both pollutants were similar to our main results for Models D. Full results of sensitivity analyses are shown in Supplemental Material Table S3.

4. Discussion

We found evidence of adverse associations between long-term exposure to $PM_{2.5}$ and NO_2 and all-cause mortality at the relatively low levels of exposure found in Sydney. The magnitude of the HR point estimates in each of our candidate models were larger than those from other cohorts with generally higher pollution concentration environments than in our study region. Although these associations in our main confounder models were not statistically significant this suggests the slope of the $PM_{2.5}$-mortality exposure-response relation may be steeper at the lower end of the exposure range and is consistent with results from work on integrated exposure response functions reported in Burnett et al. (2014) and Burnett et al. (2018). The HRs from our study were reduced and became non-significant when we adjusted for the

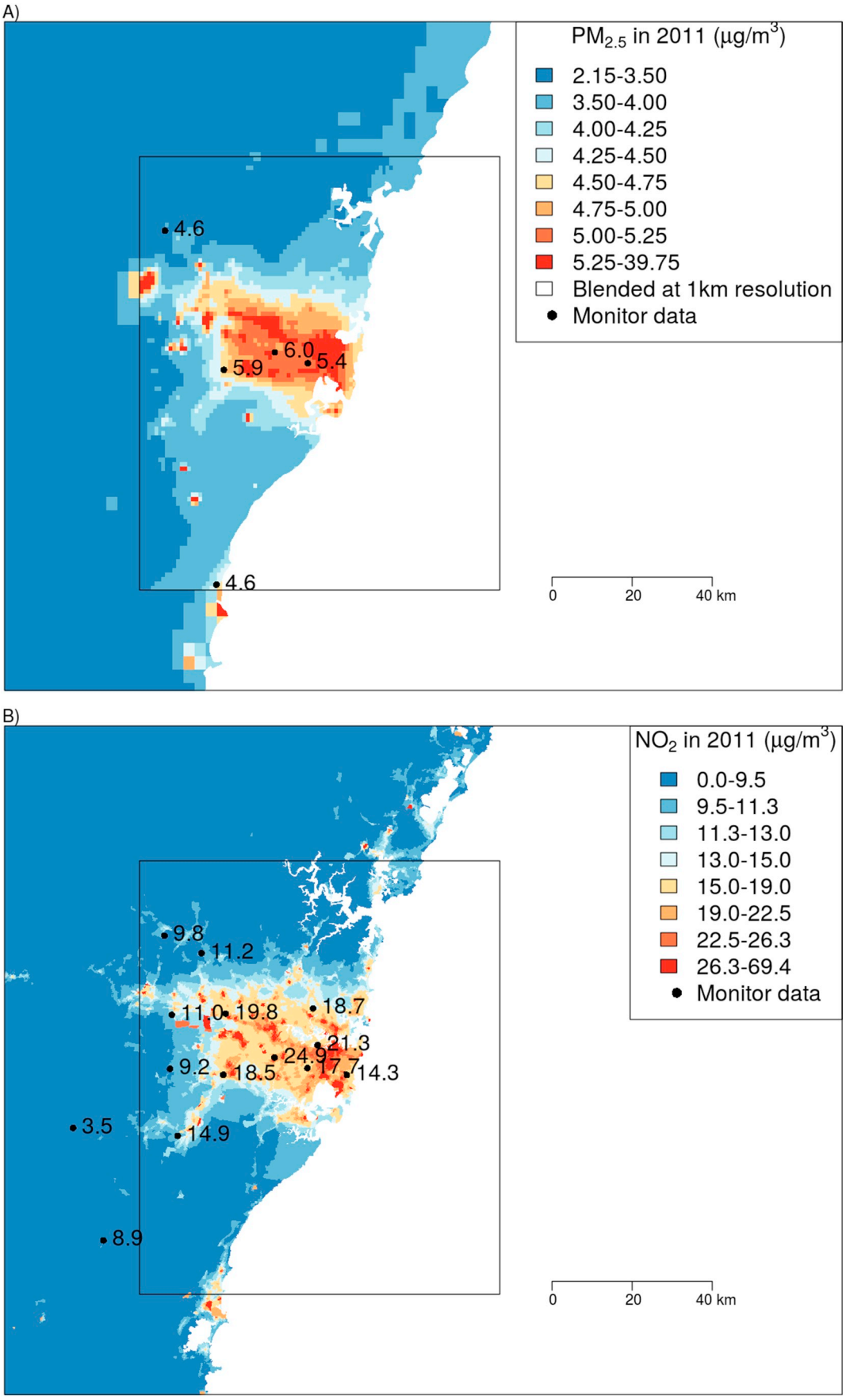

Fig. 1. Map of study area, black rectangle denotes the inner domain used for our main models. A) includes $PM_{2.5}$ annual average, and B) shows NO_2 annual average.

Table 1
Population characteristics for those study participants with baseline age < 80 years.

		N	% Of sample	No. died	% Died
Overall	Total	75,268	100.0%	3282	4.4%
Gender	Male	35,793	47.6%	2134	6.0%
	Female	39,475	52.4%	1148	2.9%
Age at baseline	45–54	27,247	36.2%	351	1.3%
	55–64	27,508	36.5%	785	2.9%
	65–74	15,616	20.7%	1230	7.9%
	75–79*	4897	6.5%	916	18.7%
Smoking	Current	5194	6.9%	400	7.7%
Status	Past	25,542	33.9%	1486	5.8%
	Never	44,532	59.2%	1396	3.1%
BMI	Underweight	880	1.2%	80	9.1%
	Normal weight	29,236	38.8%	1199	4.1%
	Overweight	29,476	39.2%	1204	4.1%
	Obese	15,676	20.8%	799	5.1%
Enrolment	2006–2007	15,614	20.7%	1057	6.8%
Year	2008–2009	59,654	79.3%	2225	3.7%
Alcoholic drinks	None	23,146	30.8%	1247	5.4%
(per week)	1–7	27,654	36.7%	999	3.6%
	8–14	13,919	18.5%	492	3.5%
	15+	10,549	14.0%	544	5.2%
Marital status	Single	17,533	23.3%	1090	6.2%
	Partnered	57,735	76.7%	2192	3.8%
Work status	Employed	44,757	59.5%	792	1.8%
	Not employed	30,511	40.5%	2490	8.2%
Education	Low (NA or School Cert.)	18,401	24.4%	1135	6.2%
	Mid (High school Cert./Trade)	31,459	41.8%	1393	4.4%
	High (University)	25,408	33.8%	754	3.0%
Born in Australia	No	26,499	35.2%	1104	4.2%
	Yes	48,769	64.8%	2178	4.5%
Sufficient physical	No	22,298	29.6%	1421	6.4%
activity	Yes	52,970	70.4%	1861	3.5%
Area SES	1st Low	16,403	21.8%	981	6.0%
	2nd	20,038	26.6%	922	4.6%
	3rd	19,656	26.1%	758	3.9%
	4th High	19,171	25.5%	621	3.2%

*Aged 80+ at baseline were excluded.

additional covariates of marital status, sufficient physical activity and area-level SES. In addition, using imputed missing values also reduced the HRs. This may reflect the small exposure contrast or bias such as exposure misclassification, confounding or overadjustment by these air pollution correlated covariates.

Our study's participants had lower exposure, on average, than those in other cohorts. The mean long-term $PM_{2.5}$ for our cohort was 4.5 μg/m^3 (range 3–14 μg/m^3) and for NO_2 was 17.8 μg/m^3 (range 8–72 μg/m^3). For comparison, the average $PM_{2.5}$ was higher (16 μg/m^3) in the follow up of the Harvard Six Cities study (Lepeule et al., 2012), and 14 μg/m^3 in the spatial analysis of American Cancer Society Cancer Prevention Study II Cohort subjects residing in California (Jerrett et al., 2013). The ESCAPE cohorts (Beelen et al., 2014) observed cohort average $PM_{2.5}$ levels ranging from 7.7 to 31.0 μg/m^3 and NO_2 from 5.2 to 59.8 μg/m^3. Cohort average NO_2 concentrations where also generally higher in the 15 cohorts included in Hoek et al. (2013) which ranged from 17.0 to 67.0 μg/m^3.

We found HRs for $PM_{2.5}$ of 1.05 per 1 μg/m^3 (95% CI 0.98–1.12) and for NO_2 of 1.03 per 5 μg/m^3 (95% CI 0.98–1.07) in our Model D which adjusted for individual-level risk factors (smoking, alcohol consumption, body-mass index, educational level, born in Australia, sufficient physical activity and working status) alongside an area-level SES index and random intercept for potential spatial clustering. Our estimated air pollution exposure levels were correlated with marital status, born in Australia and area-level SES, and these were included as they may represent important confounding factors by providing additional adjustment for unmeasured variables (such as enhanced social networks and increased cash flow for couples). When we included marital status, physical activity and area-level SES variables in Models C and D, the HRs were reduced compared with Model B, albeit still greater than one, and became non-significant at the 0.05 level. However, it is also possible that adjusting for these variables has resulted in over-adjustment (Gelman and Hill, 2007; Rothman and Greenland, 2005), reducing the magnitude and strength of associations between air pollution and mortality. We observed associations between pollution with marital status, country of birth and area-level SES. Our model checking also included testing for spatial clustering by adding location random effects and a test for a non-linear exposure-response curve. Our results for Model D were robust to these additional tests. Model checking using imputed missing data found a reduced HR for $PM_{2.5}$ of 1.02 but only slightly reduced the HR for NO_2.

In one meta-analysis of long term $PM_{2.5}$ and all-cause mortality in cohorts from Europe (Beelen et al., 2014) found a pooled effect estimate of 1.014 (95% CI 1.004–1.025) per 1 μg/m^3 increase in $PM_{2.5}$ with average cohort exposure levels between 8 and 31 μg/m^3, while another review of cohorts from North America and Europe reported a pooled estimate of 1.006 (95%CI 1.004, 1.008) per 1 μg/m^3 for all-cause mortality (average cohort exposure levels were between 4 and 28 μg/m^3)(Hoek et al., 2013). For comparison our Sydney Model D HR was larger (1.05, 95%CI 0.98–1.12) per 1 μg/m^3 and the average cohort exposure level was 4.5 μg/m^3. Regarding NO_2 (Faustini et al., 2014) reported a meta-analysis for associations with all-cause mortality from studies in Asia, North America and Europe which found similar magnitude effect estimates to our findings, with a pooled estimate of 1.08 (95% CI 1.04–1.13) per 5 μg/m^3 increase. However other meta-analyses have found lower estimates for NO_2 such as 1.005 (95% CI 0.995–1.015) per 5 μg/m^3 increase reported by (Beelen et al., 2014) and 1.027 (95% CI 1.015–1.039) per 5 μg/m^3 increase by (Hoek et al., 2013). A visual comparison is shown in Fig. 3 between our estimated HRs from Model D with 1) the HRs for $PM_{2.5}$ from the 41 cohorts reported by Burnett et al. (2018) (Panel A), and 2) the HRs for NO_2 from the 22 cohorts reported by (Beelen et al., 2014) (Panel B). We have re-scaled all HRs in Fig. 3 to represent the increased risk of an exposure increase from zero to the mean of exposure reported for each cohort. These images imply the association between $PM_{2.5}$ or NO_2 exposure with all-cause mortality may continue even at the relatively low levels in our Sydney cohort, which are among the lowest levels of the cohorts studied.

Gender differences for air pollution and mortality risk have been found in some studies but while there was some suggestion for this in our study the evidence was not strong. For example, the ESCAPE study found in men a HR for $PM_{2.5}$ of 1.03 (95% CI 1.01–1.04) per 1 μg/m^3 compared to no association for women 1.00 (95% CI 0.98–1.01) (Beelen et al., 2014). However, our results were merely suggestive that the magnitude of the effects of pollutants may be larger in men. Our models showed an effect estimate in males of 1.08 (95% CI 1.00–1.17) for $PM_{2.5}$ and 1.05 (95% CI 1.00–1.10) for NO_2; compared to null effects

Table 2
Descriptive statistics for $PM_{2.5}$ (μg/m^3) and NO_2 (μg/m^3) at participants' addresses for those with baseline age < 80 years.

Pollutant	N	Mean	Median	SD	Min–max	10th pct.	25th pct.	75th pct.	90th pct.	IQR
$PM_{2.5}$	76,938	4.49	4.47	0.61	2.78–13.81	3.74	4.10	4.92	5.22	0.82
NO_2	76,813	17.75	17.14	4.80	8.49–72.43	12.35	14.25	20.33	23.93	6.08

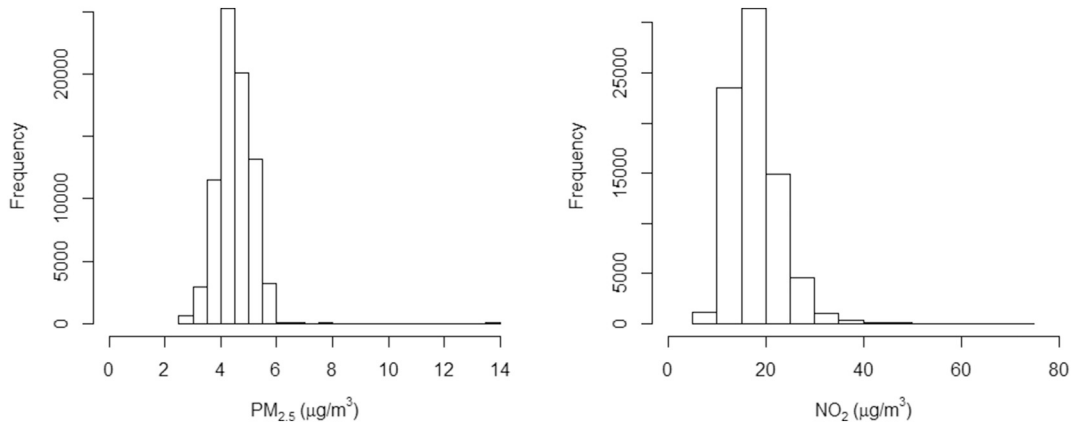

Fig. 2. Histograms for the estimated pollutant concentrations for participants baseline address.

Table 3
Hazard ratios and 95% confidence intervals for our main models of all-cause mortality and exposure to air pollution. $PM_{2.5}$ models have n = 75,268 and HR expressed per for 1 µg/m^3 change. NO_2 models have n = 75,145 and HR expressed per 5 µg/m^3 change.

Pollutant	Model	HR	95% CI
$PM_{2.5}$	Model A	1.09*	1.03–1.16
1 µg/m^3	Model B	1.08*	1.02–1.14
	Model C	1.05	0.99–1.12
	Model D	1.05	0.98–1.12
NO_2	Model A	1.04*	1.01–1.08
5 µg/m^3	Model B	1.04*	1.01–1.08
	Model C	1.03	0.99–1.07
	Model D	1.03	0.98–1.07

Model A = age as time axis, sex, and year of enrolment; Model B = smoking status, alcohol consumption, BMI, educational level, working status and born in Australia; Model C = Model B + marital status + sufficient physical activity; Model D = Model C + location random intercept + area-level SES.

* P-value ≤ 0.05.

for women ($PM_{2.5}$ = 1.01, 95% CI 0.92–1.11; NO_2 = 0.99, 0.92–1.06). However, the tests for model improvement using BIC did not show strong support for effect modification by gender.

The evidence we found supports a detrimental association between $PM_{2.5}$ and mortality. A range of plausible mechanisms for these mortality effects have been identified in the literature such as oxidative stress, respiratory irritation and toxicity of adsorbed chemical compounds (Brook et al., 2010; Franklin et al., 2015). Furthermore, our results provide some support for the suggestion that $PM_{2.5}$ may not have a 'safe' threshold and this would have important implications for policymakers implementing control strategies to reduce the health burden due to air pollution. If no safe threshold exists for $PM_{2.5}$ exposure then substantial investment in emission reduction toward background exposure levels is required, however this will be offset by substantial co-benefits to health.

We expect the air pollution and health findings from the '45 and Up Study' cohort in NSW to become increasingly important as longer periods of cohort follow up data become available in coming years (including cause specific data), along with improved estimates of both NO_2 and $PM_{2.5}$ covering all of New South Wales, not only the Sydney metropolitan region as is reported in this current study.

Our study contributes to the body of evidence required by policy makers responsible for reducing the health burden due to air pollution. There may be implications for researchers calculating cost-benefit estimates to guide changes to regulations of air pollution emissions and land use.

Our study used a large prospective cohort with high quality individual level socio-demographic information, linked mortality data and area-level SES from the population census and air pollution estimates. The large number of cohort participants resident in the single urban airshed of Sydney on the whole have experienced long-term exposure to low levels of air pollution, by international standards, and this

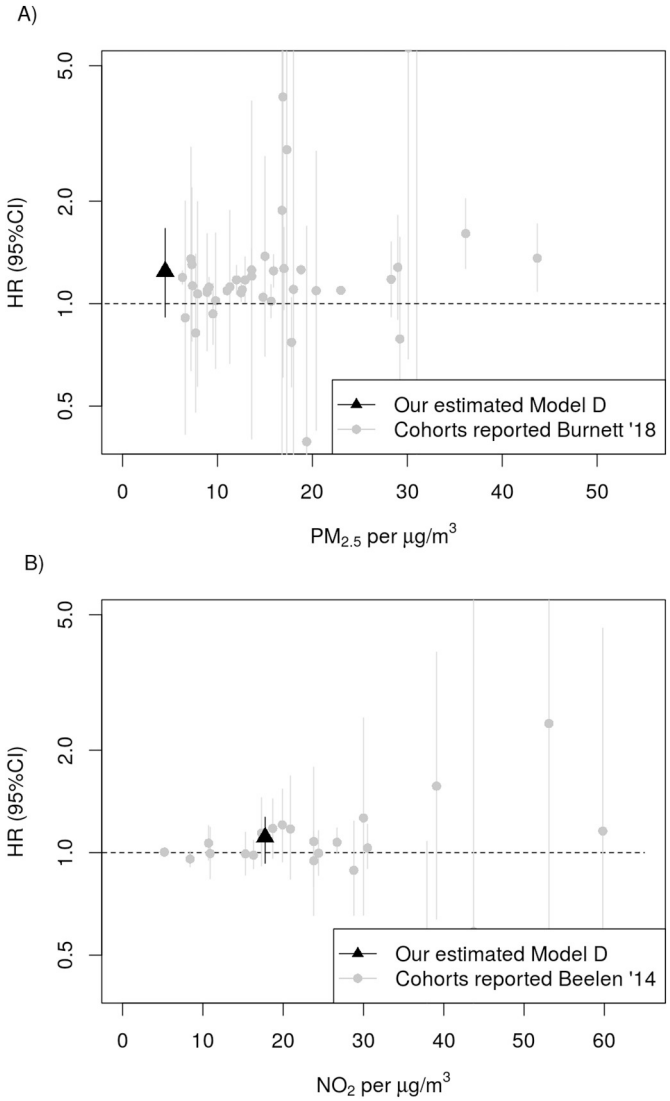

Fig. 3. Visual comparison between our estimated HRs from Model D with: A) the HRs for $PM_{2.5}$ from the 41 cohorts reported by Burnett et al. (2018), and B) the HRs for NO_2 from the 22 cohorts reported by Beelen et al. (2014).

is a key strength of our study.

Limitations of this study include the relatively short period of follow up (~7 years) and relatively large $PM_{2.5}$ spatial unit (1 km grid cells). Our exposure estimates are also likely to include error due to the lack of historical geocoded street addresses available for the previous addresses of participants, so we assumed that exposure around the time of recruitment is representative of long-term exposure.

Another limitation is due to the nature of cohort studies, including the 'healthy volunteer effect' (Lindsted et al., 1996), so that the '45 and Up Study' will not necessarily be directly representative of the general NSW population 45 and over (Banks et al., 2008). Nevertheless, methodological assessments such as Ponsonby et al. (1996) and Rothman et al. (2013) have found that associations observed within non-representative subsets of the sample population can still give reliable inferences, especially in cohort studies with good follow-up, so long as internal validity can be assumed. A key issue in assessing internal validity is potential selection bias whereby the inclusion of participant's data is conditioned on both the exposure and outcome variables in the analysis (Lash et al., 2009). This might occur in the case of systematic missing data, or if participants were more or less likely to join the study due to both their exposures and health status. While it is virtually impossible for death (our primary outcome) to influence participation (due to the linkage to the administrative death register there is minimal attrition bias), it is possible that the general health status of individuals may influence their decision to participate. Likewise, it is improbable that air pollution exposure itself would influence the likelihood of participation, however individual socio-economic variables (e.g. education) may influence an individual's decision to participate as well as their exposure. In addition, postal survey data can be less reliable than other collection methods (e.g. computer assisted telephone interviewing, CATI). We consider the likely impact of these biases to be low based on a previous analysis using the '45 and Up Study' cohort in which consistent estimates of exposure-outcome relationships were found from internal comparisons within the cohort for a broad range of risk factors when compared with a more representative population survey of NSW using CATI data collection (Mealing et al., 2010).

There are some sources of uncertainty in the CTM $PM_{2.5}$ modelled estimate even though statistics for model performance were quite good (cross-validated $R^2 = 0.70$). The estimates were generated by adding up the CTM internal $PM_{2.5}$ components from all the simulated anthropogenic and natural source groups. There is uncertainty related to some of the sources in the emissions inventory, as well as secondary aerosol production. If emission sources/secondary particle formation are not captured, therefore the CTM prediction will be an underestimate and so an additional bias correction step that adjusts the modelled output by blending it with observed data from pollution monitors was employed.

We conducted a sensitivity analysis by expanding our study region to a larger 300 km × 300 km study region. The CTM 100 km × 100 km model domain is nested within the larger 300 km × 300 km region, which is itself nested within larger regional and continental scale models. This means that local grid cells resolve information drawn from both local sources as well as regional sources. The regional information comes from national datasets so that all regional sources can contribute. Therefore, long range transport of aerosols such as sea salt, dust, fires and other upwind anthropogenic sources will be represented. With regard to our NO_2 model the regression model incorporates information drawn from all emissions sources given by a national emissions inventory within a region defined by the 10 km radius buffers around each prediction node (thus covering the entire study region as well as a buffer zone out to 10 km from the edge of the study region). We observed similar HRs to our main Model D when we used this larger 300 km × 300 km study region. We consider it likely that there may be increased exposure misclassification error, especially for $PM_{2.5}$ where the outer Sydney region only had 3 km grid cell estimates that were not blended with monitor data.

Finally, there is a potential limitation related to missing data from the death registry because it only records deaths which occurred in the state of NSW. Thus, a small number of participants who may have died in other states of Australia or overseas are not captured in this database, however numbers are likely to be so low that we consider this issue negligible for our analysis.

Our results were sensitive to the addition of the covariates: marital status, physical activity and area-level SES. Further studies on this cohort are needed to evaluate the replicability of these results and will be enhanced by improvements in exposure estimation methods and a longer period of follow-up. For example, the five-year follow-up of the 45 and Up survey was conducted between 2012 and 2015 and we will seek access to this linked dataset once the data are processed. The addition of several new years of mortality data will allow for more statistical power and cause-specific death records will enable more specificity in the analysis. Additionally, the first '45 and Up Study' cohort five-year follow-up survey will enable the potential effect of residential mobility to be included. Lastly, improved air pollution modelling will also need to be explored at higher spatial and temporal resolution, such as incorporating blending methods of ground-based monitor data with satellite images, CTM and land use data.

5. Conclusions

We found evidence of a detrimental association between $PM_{2.5}$ and NO_2 with all-cause mortality in the low pollution concentration environment in Sydney, Australia. However, a clear determination of the association with mortality is difficult because the results were not statistically significant. There was limited spatial variation in our pollution exposure metric, and the results were sensitive to some covariates such as marital status, sufficient physical activity, area-level SES and missing data imputation. We tested several different models to explore the robustness to additional parameters and functional forms. The magnitude of the hazard ratios were large compared with those reported in most other cohort studies in the literature. Those other cohorts generally came from countries which had higher pollution concentration environments than Sydney. This may imply a steeper slope of the exposure-response curve at the lower end of the spectrum and health impacts even at low levels of pollution concentration. Our results will contribute to global meta-analysis and will enable more precise effect estimates for health impact assessments.

Acknowledgements, including grant information

IH, FS and LK were supported by The Centre for Air Pollution, Energy and Health Research (CAR). CAR (www.car-cre.org.au) is an Australian National Health and Medical Research Council (NHMRC) funded Centre for Research Excellence (APP1030259). YG was supported by the Career Development Fellowship of NHMRC (#APP1107107).

This research was completed using data collected through the '45 and Up Study' (www.saxinstitute.org.au). The '45 and Up Study' is managed by the Sax Institute in collaboration with major partner Cancer Council NSW; and partners: the National Heart Foundation of Australia (NSW Division); NSW Ministry of Health; NSW Government Family & Community Services – Ageing, Carers and the Disability Council NSW; and the Australian Red Cross Blood Service. We thank the many thousands of people participating in the 45 and Up Study.

Informed consent was obtained from all survey participants. The study presented in this paper was approved by the NSW Population & Health Services Research Ethics Committee AU RED Reference: HREC/15/CIPHS/4, Cancer Institute NSW reference number: 2015/02/575 for Sub-study 2 [Air Pollution, Traffic Exposures and Mortality and Morbidity in Older Australians (APTEMA) Study], under the overarching project titled 'Understanding the impact of the social, economic and environmental factors on the health of Australians in mid - later

life; where are the opportunities for prevention?' (NHMRC grant 402,810). Thanks go to Associate Professor Philayrath Phongsavan at The University of Sydney for coordination of the data cleaning.

Competing financial interests declaration

There are no competing financial interests for the authors of this paper.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.envint.2019.02.044.

References

Apte, J.S., Marshall, J.D., Cohen, A.J., Brauer, M., 2015. Addressing global mortality from ambient $PM_{2.5}$. Environ. Sci. Technol. 49, 8057–8066.

Banks, E., Redman, S., Jorm, L., Armstrong, B., Bauman, A., Beard, J., et al., 2008. Cohort profile: the 45 and up study. Int. J. Epidemiol. 37, 941–947.

Beadle, G.F., McCarthy, N.J., Baade, P.D., 2013. Effect of age at diagnosis of breast cancer on the patterns and risk of mortality from all causes: a population-based study in Australia. Asia Pac. J. Clin. Oncol. 9, 129–138.

Beelen, R., Raaschou-Nielsen, O., Stafoggia, M., Andersen, Z.J., Weinmayr, G., Hoffmann, B., 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. Lancet 383, 785–795.

Brook, R.D., Rajagopalan, S., Pope 3rd, C.A., Brook, J.R., Bhatnagar, A., Diez-Roux, A.V., et al., 2010. Particulate matter air pollution and cardiovascular disease: an update to the scientific statement from the American Heart Association. Circulation 121, 2331–2378.

Burnett, R.T., Pope 3rd, C.A., Ezzati, M., Olives, C., Lim, S.S., Mehta, S., et al., 2014. An integrated risk function for estimating the global burden of disease attributable to ambient fine particulate matter exposure. Environ. Health Perspect. 122, 397–403.

Burnett, R., Chen, H., Szyszkowicz, M., Fann, N., Hubbell, B., Pope 3rd, C.A., et al., 2018. Global estimates of mortality associated with long-term exposure to outdoor fine particulate matter. Proc. Natl. Acad. Sci. U. S. A. 115, 9592–9597.

Cesaroni, G., Forastiere, F., Stafoggia, M., Andersen, Z.J., Badaloni, C., Beelen, R., 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.

Chen, H., Goldberg, M.S., Villeneuve, P.J., 2008. A systematic review of the relation between long-term exposure to ambient air pollution and chronic diseases. Rev. Environ. Health 23, 243–297.

Cohen, A.J., Brauer, M., Burnett, R., Anderson, H.R., Frostad, J., Estep, K., et al., 2017. Estimates and 25-year trends of the global burden of disease attributable to ambient air pollution: an analysis of data from the global burden of diseases study 2015. Lancet 389, 1907–1918.

Cope, M., Keywood, M., Emmerson, K., Galbally, I., Boast, K., Chambers, S., et al., 2014. The Sydney particle study. In: The Centre for Australian Weather and Climate Research.

Di, Q., Wang, Y., Zanobetti, A., Wang, Y., Koutrakis, P., Choirat, C., et al., 2017. Air pollution and mortality in the Medicare population. New Engl. J. Med. 376, 2513–2522.

Dockery, D.W., Pope III, A.C., Xu, X., Spengler, J.D., Ware, J.H., Fay, M.E., et al., 1993. An association between air pollution and mortality in six U.S. cities. New Engl. J. Med. 329, 1753–1759.

ESCAPE Statistical Tasks Working Group, 2012. Study Protocol Escape Statistical Tasks Working Group, version 7.

Faustini, A., Rapp, R., Forastiere, F., 2014. Nitrogen dioxide and mortality: review and meta-analysis of long-term studies. Eur. Respir. J. 44, 744–753.

Forouzanfar, M.H., Afshin, A., Alexander, L.T., Anderson, H.R., Bhutta, Z.A., Biryukov, S., et al., 2016. Global, regional, and national comparative risk assessment of 79 behavioural, environmental and occupational, and metabolic risks or clusters of risks, 1990–2015: a systematic analysis for the global burden of disease study 2015. Lancet 388, 1659–1724.

Franklin BA, Brook R, Arden Pope C, 3rd. 2015. Air pollution and cardiovascular disease. Curr. Probl. Cardiol. 40:207–238.

Gelman, A., Hill, J., 2007. Do not control for post-treatment variables. In: Data Analysis Using Regression and Multilevel/Hierarchical Models. Cambridge, United Kingdom. Cambridge University Press, pp. 188–190.

Hanley, J.A., Foster, B.J., 2014. Avoiding blunders involving 'immortal time'. Int. J. Epidemiol. 43, 949–961.

Hoek, G., Krishnan, R.M., Beelen, R., Peters, A., Ostro, B., Brunekreef, B., et al., 2013. Long-term air pollution exposure and cardio- respiratory mortality: a review. Environ. Health 12, 43.

Jerrett, M., Burnett, R.T., Beckerman, B.S., Turner, M.C., Krewski, D., Thurston, G., et al., 2013. Spatial analysis of air pollution and mortality in California. Am. J. Respir. Crit. Care Med. 188, 593–599.

Knibbs, L.D., Hewson, M.G., Bechle, M.J., Marshall, J.D., Barnett, A.G., 2014. A national satellite-based land-use regression model for air pollution exposure assessment in Australia. Environ. Res. 135, 204–211.

Knibbs, L.D., Coorey, C.P., Bechle, M.J., Cowie, C.T., Dirgawati, M., Heyworth, J.S., et al., 2016. Independent validation of national satellite-based land-use regression models for nitrogen dioxide using passive samplers. Environ. Sci. Technol. 50, 12331–12338.

Lash, T.L., Fox, M.P., Fink, AK., 2009. Applying quantitative bias analysis to epidemiologic data: Springer-Verlag New York.

Lepeule, J., Laden, F., Dockery, D., Schwartz, J., 2012. Chronic exposure to fine particles and mortality: an extended follow-up of the Harvard six cities study from 1974 to 2009. Environ. Health Perspect. 120, 965–970.

Lindsted, K.D., Fraser, G.E., Steinkohl, M., Beeson, W.L., 1996. Healthy volunteer effect in a cohort study: temporal resolution in the Adventist health study. J. Clin. Epidemiol. 49, 783–790.

Mealing, N.M., Banks, E., Jorm, L.R., Steel, D.G., Clements, M.S., Rogers, K.D., 2010. Investigation of relative risk estimates from studies of the same population with contrasting response rates and designs. BMC Med. Res. Methodol. 10, 26.

NSW Office of Environment and Heritage. n.d. Air quality data. Retrieved from http://www.environment.nsw.gov.au/AQMS/search.htm.

Physick, W.L., Cope, M.E., Lee, S., Hurley, P.J., 2007. An approach for estimating exposure to ambient concentrations. J Expo Sci Environ Epidemiol 17, 76–83.

Pinault, L., Tjepkema, M., Crouse, D.L., Weichenthal, S., van Donkelaar, A., Martin, R.V., et al., 2016. Risk estimates of mortality attributed to low concentrations of ambient fine particulate matter in the Canadian community health survey cohort. Environ. Health 15, 18.

Ponsonby, A.-L., Dwyer, T., Couper, D., 1996. Is this finding relevant? Generalisation and epidemiology. Aust. N. Z. J. Public Health 20, 54–56.

Pope III, C.A., Thun, M.J., Namboodiri, M.M., Dockery, D.W., Evans, J.S., Speizer, F.E., et al., 1995. Particulate air pollution as a predictor of mortality in a prospective study of US adults. Am. J. Respir. Crit. Care Med. 151, 669–674.

R Core Team, 2017. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.

Raaschou-Nielsen, O., Andersen, Z.J., Beelen, R., Samoli, E., Stafoggia, M., Weinmayr, G., 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.

Raftery, A.E., 1995. Bayesian model selection in social research. Sociol. Methodol. 25, 111–163.

Rothman, K.J., Greenland, S., 2005. Causation and causal inference in epidemiology. Am. J. Public Health 95 (Suppl. 1), S144–S150.

Rothman, K.J., Gallacher, J.E., Hatch, E.E., 2013. Why representativeness should be avoided. Int. J. Epidemiol. 42, 1012–1014.

Shariff, S.Z., Cuerden, M.S., Jain, A.K., Garg, A.X., 2008. The secret of immortal time bias in epidemiologic studies. J. Am. Soc. Nephrol. 19, 841–843.

Shmueli, G., 2010. To explain or to predict? Stat. Sci. 25, 289–310.