BMJ Open

Long-term ambient air pollution exposure and self-reported morbidity in the Australian Longitudinal Study on Women's Health: a cross-sectional study

Nina Lazarevic,[1] Annette J Dobson,[1] Adrian G Barnett,[2] Luke D Knibbs[1]

To cite: Lazarevic N, Dobson AJ, Barnett AG, *et al*. Long-term ambient air pollution exposure and self-reported morbidity in the Australian Longitudinal Study on Women's Health: a cross-sectional study. *BMJ Open* 2015;**5**:e008714. doi:10.1136/bmjopen-2015-008714

▸ Prepublication history and additional material is available. To view please visit the journal (http://dx.doi.org/10.1136/bmjopen-2015-008714).

Received 13 May 2015
Revised 8 September 2015
Accepted 24 September 2015

[1]School of Public Health, Faculty of Medicine and Biomedical Sciences, The University of Queensland, Brisbane, Queensland, Australia
[2]School of Public Health and Social Work, Faculty of Health, Queensland University of Technology, Brisbane, Queensland, Australia

Correspondence to
Dr Luke D Knibbs;
l.knibbs@uq.edu.au

ABSTRACT

Objective: We sought to assess the effect of long-term exposure to ambient air pollution on the prevalence of self-reported health outcomes in Australian women.

Design: Cross-sectional study.

Setting and participants: The geocoded residential addresses of 26 991 women across 3 age cohorts in the Australian Longitudinal Study on Women's Health between 2006 and 2011 were linked to nitrogen dioxide (NO_2) exposure estimates from a land-use regression model. Annual average NO_2 concentrations and residential proximity to roads were used as proxies of exposure to ambient air pollution.

Outcome measures: Self-reported disease presence for diabetes mellitus, heart disease, hypertension, stroke, asthma, chronic obstructive pulmonary disease and self-reported symptoms of allergies, breathing difficulties, chest pain and palpitations.

Methods: Disease prevalence was modelled by population-averaged Poisson regression models estimated by generalised estimating equations. Associations between symptoms and ambient air pollution were modelled by multilevel mixed logistic regression. Spatial clustering was accounted for at the postcode level.

Results: No associations were observed between any of the outcome and exposure variables considered at the 1% significance level after adjusting for known risk factors and confounders.

Conclusions: Long-term exposure to ambient air pollution was not associated with self-reported disease prevalence in Australian women. The observed results may have been due to exposure and outcome misclassification, lack of power to detect weak associations or an actual absence of associations with self-reported outcomes at the relatively low annual average air pollution exposure levels across Australia.

Strengths and limitations of this study

- This is the first Australian national-scale air pollution study examining the effects of long-term ambient air pollution exposure on chronic morbidity in women.
- Individual-level data on numerous self-reported medical conditions and confounding variables were linked to exposure estimates at residential addresses using a validated national land-use regression model.
- An inherently low signal-to-noise ratio with uncertainty in both outcome and exposure measurement may have biased the results towards the null.

INTRODUCTION

Ambient air pollution is one of the leading environmental risk factors in the global burden of disease.[1] Current evidence suggests adverse effects of air pollution on cardiovascular, respiratory, metabolic and allergic diseases, with outdoor particulate matter (PM) air pollution responsible for approximately 3% of global cardiorespiratory mortality.[2]

Although air pollution levels in Australia are considered low in comparison to other economically developed nations,[3] the Australian population is concentrated in major cities where air pollution exposure is ubiquitous and more likely to be elevated due to proximity with emissions sources. Given the large proportion of the population exposed to air pollution, even small estimated effects would increase the risks of air pollution-associated morbidity and mortality in Australia.

Recent systematic reviews and meta-analyses suggest modest positive associations between ambient air pollution and cardiorespiratory and metabolic diseases.[4–7] Short-term air pollution effects have been extensively studied, especially in relation to cardiorespiratory outcomes. Long-term effects of cumulative exposure at ambient levels are less understood despite concern that chronic exposure increases morbidity and mortality risk to a greater extent than

short-term exposure.[8] In the case of respiratory and allergic diseases, air pollution is known to result in acute exacerbations of existing conditions, but there is uncertainty in the role of air pollution in the development of adult-onset disease.[9 10]

Differential effects of air pollution exposure by gender have been recorded, with evidence of stronger associations in women.[11–13] Differences in male and female lung architecture and the effects of hormonal status have been proposed as explanations[14]; however, few studies have limited their attention to women specifically. The aim of this study was to determine if there are associations between ambient air pollution and the prevalence of several chronic health conditions among Australian women. We thus sought to add to the limited evidence base on the specific effects of long-term exposure to ambient air pollution on women in a relatively low pollution setting.

METHODS

Australian Longitudinal Study on Women's Health

The Australian Longitudinal Study on Women's Health (ALSWH) is a population-based prospective longitudinal study that started in 1996 to assess factors that affect the health of Australian women. Participants were recruited randomly from the Medicare database (Australia's universal healthcare scheme), with deliberate overrepresentation of women living in non-urban areas to account for the marked concentration of the Australian population in coastal cities.[15] Participants are surveyed by mail every 3 years to collect a self-reported assessment of their physical and emotional health, health-related behaviours, risk factors, and sociodemographic characteristics. The ALSWH is approved by the research ethics committees of the University of Queensland and the University of Newcastle. Further details of the study can be found online (http://www.alswh.org.au) and in Lee *et al*.[15]

The present study focuses on survey responses from 26 991 participants collected in the fourth (2006) and fifth surveys (2009) of women born between 1973 and 1978 (the 'younger' cohort, aged 31–36 years at the later survey), the fifth (2007) and sixth (2010) surveys of women born between 1946 and 1951 (the 'middle-aged' cohort, aged 59–64 years at the later survey), and the fifth (2008) and sixth (2011) surveys of women born between 1921 and 1926 (the 'older' cohort, aged 85–90 years at the later survey). This time frame was selected to match the availability of exposure data. Only de-identified data were used for privacy reasons; address geocoding and exposure assignment were conducted separately to data analysis, with only alias postcodes available to the data analyst.

Outcome measures

The study examined the self-reported presence of six diseases with plausible links to air pollution: diabetes mellitus, heart disease (includes angina, heart attack, other heart-related problems), hypertension, stroke, asthma, chronic obstructive pulmonary disease (COPD; includes bronchitis and emphysema). We also examined self-reported symptoms of allergies (includes hay fever and sinusitis), breathing difficulties, chest pain, and palpitations. We used iron deficiency as a negative control[16] as it is not believed to be associated with ambient air pollution, but is likely to be affected by the same unobserved confounders as the outcome–exposure relationships of interest.

Disease data were obtained from survey questions asking participants whether they had been diagnosed or treated for the medical condition in question in the previous 3 years. Responses were dichotomous. Symptom data were obtained from survey questions asking participants how often they had experienced a particular symptom in the previous 12 months. Responses were on a four-point ordinal scale: never, rarely, sometimes or often. The symptom frequency response variable was dichotomised to avoid the subjectivity associated with the given ordinal scale and reporting heterogeneity bias.[17 18]

Exposure data

Ambient air pollution exposures were estimated using a national satellite-based land-use regression model, described in detail elsewhere.[19] Briefly, it is capable of capturing 81% of spatial variability in annual mean ambient NO_2 levels across Australia, with a prediction error of 19%. We used NO_2 as a proxy for ambient air pollution because it exhibits greater spatial heterogeneity than other ambient air pollutants, and is produced by major ambient pollution sources like motor vehicles and industrial processes.[20] Exposure estimates were assigned to geocoded residential addresses, with long-term ambient air pollution exposure defined as the predicted annual mean NO_2 concentration at the place of residence. We did not have access to comparable national-scale models for other pollutants.

We also assessed residential proximity to major and minor roads at the time of survey completion as a proxy for traffic-related ambient air pollution. Major and minor road definitions were based on the Public Sector Mapping Agencies Australia Limited road classification hierarchy.[19] Where an address could not be matched exactly during geocoding, the next best match was attempted (eg, next door or same street), and if no matching street number or name could be located, matches were made to the postcode centroid. We assessed the sensitivity of our results to the accuracy of address geocoding.

For the disease prevalence analysis, a 3-year average annual NO_2 exposure was assigned (over the year of the survey and 2 years prior) to correspond to the 3-year time period in the survey outcome questions. For the symptom prevalence analysis, the annual mean NO_2 concentration in the year immediately prior to the survey year was assigned as the exposure. This was

selected because surveys were conducted at the beginning of the calendar year and variability in annual mean NO_2 is negligible in proximate years (correlation coefficients between the years of the study period were all greater than 0.99).

Sample selection

The disease questions were asked across all three age cohorts except in the case of stroke, which was asked only in surveys of middle-aged and older women. Only the second of the two surveys for each respondent (between 2009 and 2011) were used because exposure estimates were only available between 2006 and 2011, and exposures were required to correspond to the 3-year period referred to in the disease questions. The disease prevalence analysis was thus performed on pooled data from the fifth survey of the younger cohort, and the sixth surveys of the middle-aged and older cohorts (see online supplementary section 1 for further details), except for the stroke outcome for which data on the younger cohort were unavailable.

Responses from both surveys of respondents in each age cohort were included in the symptom analysis. Allergy and palpitation outcomes were only assessed in surveys of younger and middle-aged women; chest pain was assessed in surveys of middle-aged and older women; and breathing difficulty was assessed across all three cohorts, but not in survey 4 of younger women.

We restricted the sample to those respondents with no missing data on the variables used in the analysis to first exclude respondents with missing outcome, exposure and postcode data. We then restricted the disease prevalence analysis and models which accounted for clustering by postcode to movement below an arbitrarily chosen 5 km residential mobility threshold (see online supplementary table S1). We did this to avoid unduly excluding participants whose apparent movement over short distances between surveys was more likely due to differences in geocoding results rather than an actual change of residence (we did not have access to residential addresses for comparison). At the same time, we wanted to identify participants who had moved during the preceding 3-year period and were not suitable for the disease prevalence analyses. We assessed the sensitivity of our results to this choice of mobility threshold (see online supplementary section 2 for further details).

Response rates in the ALSWH for the younger cohort surveys were 71.1% for survey 4 and 61.4% for survey 5; in the middle-aged cohort, these were 86% for survey 5 and 83% for survey 6; in the older cohort these rates were 77.4% for survey 5 and 70% for survey 6. Table 1 shows sample composition by survey and cohort, and online supplementary table S1 shows the derivation of analytical sample sizes with the number of missing observations by outcome.

Covariates included as confounders

Covariates included for confounding control included age group, body mass index (BMI), smoking status, alcohol intake, physical activity, fruit and vegetable consumption, degree of residential urbanisation or remoteness, annual mean temperature, marital status, educational attainment and self-assessed financial resources. The definitions used for these covariates are given in online supplementary section 3. We selected these confounders on the basis that they may be associated with the outcomes of interest and with exposure. Online supplementary tables S2 and S3 show descriptive statistics by outcome and covariate for the disease and symptom data, respectively.

Statistical analysis

Spatial autocorrelation

To partially account for the spatial autocorrelation observed in the exposure data, respondents were assumed to be clustered within postcodes. Alias postcodes were used for privacy reasons, which have a one-to-one mapping with actual postcodes. Latitude and longitude data were not available to the analyst due to confidentiality restrictions which precluded specification of a model that accounts for spatial autocorrelation more precisely. Incorporating clustering by postcode assumes non-zero correlation between women living within the same postcode area, but not for women living in different postcodes, regardless of their actual spatial distance. Women residing in neighbouring postcodes were assumed independent, even if they were spatially proximal.

Prevalence of self-reported disease

Relative risks (risk ratios, RRs) were modelled via Poisson regression with a log link function and robust error variance, which is known to produce consistent

Table 1 Number of respondents across survey and cohort

	Cohort			
Response to	Younger	Middle-aged	Older	Total
Only survey 4	1865 (18.5%)			1865 (6.9%)
Only survey 5	920 (9.1%)	1112 (10.0%)	1748 (30.1%)	3780 (14.0%)
Only survey 6		485 (4.4%)	243 (4.2%)	728 (2.7%)
Both surveys	7280 (72.3%)	9526 (85.6%)	3812 (65.7%)	20618 (76.4%)
Total	10065 (100%)	11123 (100%)	5803 (100%)	26991 (100%)

and efficient estimates of relative risk with binary data.[21 22] The model chosen was a population-averaged model estimated by generalised estimating equations assuming an exchangeable correlation structure using the function xtgee in Stata 13.1 (StataCorp, College Station, Texas, USA). A linear association was assumed for the NO_2 exposure and RR estimates are for an IQR increase. Proximities to major and minor roads were log-transformed (base 2) as the distribution was concentrated near zero and highly right-skewed; thus, estimated coefficients refer to a doubling in the exposure distance.

Prevalence of self-reported symptoms

ORs were estimated via a two-level mixed logistic regression with a random intercept that accounts for the dependence between repeated observations for each respondent. To partially account for spatial autocorrelation, an additional three-level mixed logistic regression with random intercepts for both respondent and postcode was used that accounts for the dependence between observations at both levels (see online supplementary section 4 for further details). Random intercepts were assumed to be normally distributed. Models were estimated by mean-variance adaptive Gauss-Hermite quadrature using the Stata 13.1 function melogit. Preliminary models were estimated with 10 quadrature points and the final model was estimated with 10, 15 and 20 quadrature points to check the stability of the results.

Models estimated

We first estimated crude effect estimates and then second adjusted for confounders and known risk factors. Known risk factors for each outcome were included as a minimum: specifically, age cohort and smoking status were controlled in all models.[23] Other confounders were identified by the 'change-in-estimate' rule of thumb;[24 25] we required only a relatively small 5% change as effect estimates were expected to be small. Physical activity and BMI group were always included in models of diabetes and cardiovascular outcomes. BMI group was always included in the asthma model. Alcohol use was always included in models of hypertension and stroke. Comorbidities were not considered.

To assess the sensitivity of our results to the accuracy of address geocoding, we re-estimated the adjusted model on the subset of respondents with exact matches between respondent addresses at survey and the geocoded national address reference file. Finally, to assess residual confounding by variables excluded by this procedure, a further model containing all available covariates was estimated (here termed as a 'fully adjusted' model).

Model comparison was assisted by the quasi-likelihood under-the-independence-model information criterion[26 27] in the models of disease prevalence and Akaike's information criterion (AIC) in the models of symptom prevalence. Details of goodness-of-fit assessment are provided in online supplementary section 5. Power analyses were conducted by simulation to determine the minimum detectable effect sizes given our model and sample size. CIs at the 99% level were used to reduce the risk of type 1 errors.

RESULTS

Exposures

Figure 1 shows distributions of annual mean NO_2 levels and distance to roads at survey 5 of the younger cohort and surveys 6 of the middle-aged and older cohorts. A high level of spatial autocorrelation was observed in the NO_2 data using the Moran's I statistic (I=0.89, Z=357.1, p<0.001).

Prevalence of self-reported disease

Table 2 shows crude and adjusted relative risk estimates for the disease outcomes. Most adjusted RR estimates were close to unity for the NO_2 exposure, and the 99% CIs included unity for all disease outcomes considered after adjusting for risk factors and confounders. Similarly, there was no evidence of association between any of the outcomes considered and residential proximity to major or minor roads. This was also the case when restricting the sample to those respondents with exact matches between address recorded at survey and the geocoding reference data. Adjusting for all available covariates did not suggest any residual confounding by the additional available variables and minimal change in precision of the estimated effects. Sensitivity analyses of the 5 km threshold for residential mobility suggested that the results were not affected by moving the threshold above or below 5 km (see online supplementary figures S1 and S2). Furthermore, no associations were observed with iron deficiency (as a negative control),

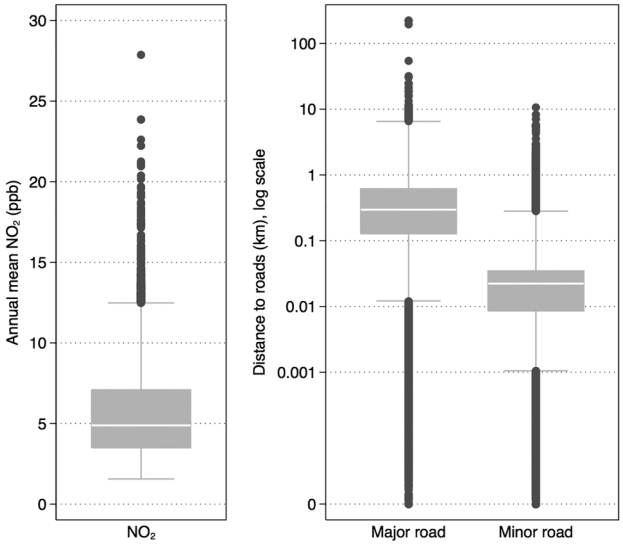

Figure 1 Box plots of annual mean NO_2 levels and distance to roads at survey 5 of the younger cohort and surveys 6 of the middle-aged and older cohorts.

Table 2 Relative risk of self-reported disease with an IQR increase in the 3-year mean NO_2 concentration or a doubling in the distance to a major or minor road

	Crude model		Adjusted model*		Fully adjusted model†		Exact geocoding subsample*	
	N	RR (99% CI)	N	RR (99% CI)	N	RR (99% CI)	N	RR (99% CI)
Diabetes								
3-year mean annual NO_2 (3.7 ppb)	14563	0.79 (0.70 to 0.89)	12443	1.04 (0.91 to 1.20)	12177	1.04 (0.90 to 1.20)	9738	1.00 (0.85 to 1.18)
Distance to major road (doubling)	14563	1.04 (1.00 to 1.07)	12940	0.99 (0.95 to 1.04)	12177	0.99 (0.95 to 1.04)	10103	0.99 (0.93 to 1.05)
Distance to minor road (doubling)	14563	1.11 (1.07 to 1.15)	12940	0.99 (0.95 to 1.04)	12177	0.98 (0.94 to 1.02)	10103	0.94 (0.87 to 1.02)
Heart disease								
3-year mean annual NO_2 (3.7 ppb)	14563	0.85 (0.77 to 0.93)	12452	0.94 (0.85 to 1.04)	12177	0.88 (0.76 to 1.01)	9739	0.90 (0.80 to 1.00)
Distance to major road (doubling)	14563	1.03 (1.00 to 1.05)	12940	1.01 (0.97 to 1.05)	12177	1.01 (0.97 to 1.05)	10103	1.01 (0.96 to 1.06)
Distance to minor road (doubling)	14563	1.11 (1.08 to 1.15)	12940	0.98 (0.95 to 1.02)	12177	0.98 (0.94 to 1.01)	10103	1.01 (0.93 to 1.09)
Hypertension								
3-year mean annual NO_2 (3.7 ppb)	14563	0.83 (0.78 to 0.87)	12395	0.97 (0.92 to 1.01)	12177	0.99 (0.94 to 1.05)	9687	0.96 (0.91 to 1.01)
Distance to major road (doubling)	14563	1.03 (1.02 to 1.05)	12880	1.00 (0.98 to 1.02)	12177	1.00 (0.98 to 1.02)	10048	1.00 (0.98 to 1.02)
Distance to minor road (doubling)	14563	1.12 (1.10 to 1.14)	12880	1.01 (0.99 to 1.02)	12177	1.00 (0.99 to 1.02)	10048	1.01 (0.98 to 1.04)
Stroke								
3-year mean annual NO_2 (3.3 ppb)	10402	0.99 (0.80 to 1.22)	8518	0.83 (0.58 to 1.19)	8384	0.74 (0.51 to 1.09)	6916	0.73 (0.49 to 1.09)
Distance to major road (doubling)	10402	0.96 (0.89 to 1.05)	8964	1.01 (0.90 to 1.14)	8384	1.02 (0.90 to 1.16)	7252	1.06 (0.91 to 1.24)
Distance to minor road (doubling)	10402	0.98 (0.89 to 1.08)	8964	0.98 (0.88 to 1.10)	8384	0.99 (0.88 to 1.11)	7252	1.13 (0.94 to 1.35)
Asthma								
3-year mean annual NO_2 (3.7 ppb)	14563	0.96 (0.89 to 1.05)	13660	0.99 (0.91 to 1.08)	12177	0.97 (0.86 to 1.10)	10658	0.95 (0.86 to 1.06)
Distance to major road (doubling)	14563	1.00 (0.97 to 1.03)	13660	1.00 (0.98 to 1.03)	12177	1.00 (0.97 to 1.03)	10658	1.01 (0.97 to 1.05)
Distance to minor road (doubling)	14563	1.00 (0.98 to 1.03)	13660	1.00 (0.97 to 1.03)	12177	1.01 (0.98 to 1.04)	10658	1.01 (0.95 to 1.07)
COPD								
3-year mean annual NO_2 (3.7 ppb)	14563	1.02 (0.92 to 1.13)	14480	0.96 (0.83 to 1.09)	12177	1.00 (0.86 to 1.17)	11294	0.92 (0.78 to 1.09)
Distance to major road (doubling)	14563	1.01 (0.97 to 1.04)	14489	1.01 (0.97 to 1.04)	12177	1.01 (0.98 to 1.05)	11295	1.00 (0.95 to 1.06)
Distance to minor road (doubling)	14563	0.99 (0.96 to 1.02)	14489	0.99 (0.95 to 1.02)	12177	1.00 (0.95 to 1.04)	11295	0.96 (0.89 to 1.04)
Iron deficiency—negative control								
3-year mean annual NO_2 (3.7 ppb)	14563	1.17 (1.09 to 1.26)	14563	1.04 (0.96 to 1.12)	12177	1.04 (0.93 to 1.16)	11355	1.02 (0.93 to 1.12)
Distance to major road (doubling)	14563	0.95 (0.93 to 0.97)	14563	0.98 (0.96 to 1.00)	12177	0.99 (0.97 to 1.01)	11355	0.98 (0.94 to 1.02)
Distance to minor road (doubling)	14563	0.93 (0.91 to 0.95)	14563	1.00 (0.98 to 1.03)	12177	1.01 (0.98 to 1.04)	11355	1.01 (0.95 to 1.07)

Sample size (N) and RRs with 99% CIs in parentheses.
*Adjusted for known risk factors and identified confounders, specifically: diabetes models adjusted for cohort, BMI group, smoking status, physical activity, and additionally highest qualification and residential remoteness in the NO_2 model; heart disease models adjusted for cohort, BMI group, smoking status, physical activity and additionally highest qualification in the NO_2 model; hypertension models adjusted for cohort, BMI group, smoking status, physical activity, alcohol use and additionally highest qualification in the NO_2 models; stroke models adjusted for cohort, BMI group, smoking status, physical activity, alcohol use, and additionally highest qualification, marital status, and residential remoteness in the NO_2 model; asthma models adjusted for cohort, smoking status and BMI group; COPD models adjusted for cohort and smoking status, and additionally residential remoteness in the NO_2 model; iron deficiency models adjusted for cohort only.
†Adjusted for all available covariates.
BMI, body mass index; COPD, chronic obstructive pulmonary disease; RR, risk ratio.

suggesting our results were not affected by unobserved confounding. Power simulations suggested that the models of stroke prevalence and most NO_2 models may not have sufficient power to detect weak associations close to unity (see online supplementary tables S4 and S5).

Prevalence of self-reported symptoms

Table 3 shows crude and adjusted ORs for the symptom outcomes. Again there was no evidence of association between annual mean NO_2 levels, or residential proximity to major or minor roads, and any of the symptoms considered after adjusting for risk factors and confounders. This was also the case when adjusting for all available covariates, when the analytical sample was restricted to those respondents with exact matches between address recorded at survey and the geocoding reference data, and when three-level models were estimated which account for clustering by postcode. The findings were stable when using 10, 15 or 20 integration points for the adaptive Gauss-Hermite quadrature.

Spatial autocorrelation

The estimated within-postcode correlations in the disease prevalence models were negligible (<0.012 in absolute value in all cases), suggesting little benefit in estimating this additional parameter. Within-postcode correlations were also of negligible magnitude in the symptom prevalence models (≤0.01 in absolute value in all cases), and there was negligible difference in AIC when comparing the three-level models with their corresponding nested two-level models, suggesting no clustering by postcode.

DISCUSSION

This is the first Australian national-scale air pollution study focusing on the effects of long-term ambient air pollution exposure on chronic morbidity in women. We linked individual-level data on self-reported symptoms and diseases (and confounding variables) to residential exposure estimates from a validated land-use regression model and accounted for spatial clustering of respondents. We found no evidence of associations between self-reported diabetes, heart disease, hypertension, stroke, asthma, COPD or self-reported symptoms of allergies, breathing difficulties, chest pain or palpitations and ambient NO_2 air pollution exposure or residential proximity to roads after adjusting for known risk factors and confounders.

Comparison with other studies

Diabetes mellitus type 2

Although experimental evidence suggests a role for air pollution in the aetiology of diabetes, epidemiological evidence linking air pollution and diabetes prevalence is limited and mixed.[28] Our results are consistent with one of two recent systematic reviews and meta-analyses of cross-sectional studies: Janghorbani *et al*[29] found no association while Balti *et al*[4] suggested exposure to ambient NO_2 air pollution may be associated with a modest increase in diabetes prevalence. Both reviews relied on a different set of two out of three available cross-sectional studies, of which two found no association[11 30] and the third observed a positive association.[28] We were unable to exclude participants with type 1 diabetes in the middle-aged and older cohorts, which may have diluted our observed effects. However, the overwhelming majority of cases in these age groups in our data set are type 2 diabetes.

Respiratory and allergic outcomes

It is well known that air pollution exacerbates symptoms in individuals already suffering from asthma and allergic conditions;[9 10 31 32] however, the role that air pollution plays in the development of asthma and allergies is less clear.[9 10] Recent systematic reviews and meta-analyses suggest an association between increases in ambient NO_2 air pollution and asthma incidence,[5] but not community-level prevalence.[33] Few studies of within-community prevalence have looked at adult asthma,[32] with some of those conducted to date reporting positive associations[34 35] and others, consistent with our study, reporting no association.[36 37] Our findings are also supported by a recent Tasmanian study that found no associations between adult asthma prevalence and traffic-related pollution.[38] In the case of allergic disease, our results are consistent with those of Pujades-Rodríguez *et al*,[37] but not with results of Cesaroni *et al*[36] who observed an association between rhinitis prevalence and various traffic-related indicators or Lindgren *et al*,[35] who observed associations between allergic rhinitis (hay fever), proximity to traffic and exposure to nitrogen oxides (NO_X). A limitation of our analysis was the lack of information with which to separate allergic from non-allergic asthma,[10] and the broad categorisation of allergies, hay fever and sinusitis as one outcome when the underlying conditions may have varying subtypes and aetiologies.

The evidence for long-term air pollution effects on COPD prevalence is also not conclusive, despite the existence of biologically plausible mechanisms and well-established evidence that air pollution affects lung function and exacerbates pre-existing COPD.[39 40] Our results are consistent with two recent meta-analyses: associations between NO_2 exposure and chronic bronchitis symptoms were not observed in the European Study of Cohorts for Air Pollution Effects[41] or in a systematic review of the effect of outdoor PM air pollution on COPD prevalence.[42] Moreover, the majority of cross-sectional studies reviewed by Schikowski *et al*[39] observed no association. In contrast, NO_2 and PM exposure were both associated with a higher risk of COPD in a study of German women,[43] while proximity to busy roads and long-term NO_X exposure were associated with a higher risk of self-reported COPD and chronic bronchitis

Table 3 ORs of self-reported symptoms with an IQR increase in the annual mean NO_2 concentration or a doubling in the distance to a major or minor road

	Crude model		Adjusted model*		Fully adjusted model†		Exact geocoding subsample*		Adjusted 3-level model*	
	N	OR (99% CI)	N	OR (99% CI)	N	OR (99% CI)	N	OR (99% CI)	N	OR (99% CI)
Allergies										
Annual mean NO_2 (3.7 ppb)	35797	1.02 (0.92 to 1.13)	35496	0.87 (0.76 to 1.01)	30410	0.87 (0.76 to 1.01)	23821	0.89 (0.77 to 1.04)	23268	0.85 (0.73 to 1.00)
Distance to major road (doubling)	35797	1.00 (0.97 to 1.03)	35676	1.01 (0.98 to 1.04)	30410	1.02 (0.98 to 1.05)	23834	1.01 (0.97 to 1.06)	23288	1.01 (0.97 to 1.04)
Distance to minor road (doubling)	35797	0.98 (0.95 to 1.01)	35676	1.01 (0.98 to 1.04)	30410	1.02 (0.98 to 1.05)	23834	0.96 (0.91 to 1.01)	23288	1.01 (0.97 to 1.06)
Breathing difficulty										
Annual mean NO_2 (3.4 ppb)	35457	0.94 (0.87 to 1.02)	31202	1.05 (0.96 to 1.15)	27257	1.09 (0.97 to 1.23)	21252	1.02 (0.92 to 1.13)	23760	1.05 (0.95 to 1.16)
Distance to major road (doubling)	35457	1.04 (1.01 to 1.06)	33618	1.00 (0.98 to 1.03)	27257	1.01 (0.98 to 1.04)	22857	1.01 (0.97 to 1.06)	24814	1.02 (0.98 to 1.05)
Distance to minor road (doubling)	35457	1.08 (1.05 to 1.10)	33618	0.99 (0.97 to 1.02)	27257	1.00 (0.97 to 1.03)	22857	0.97 (0.92 to 1.02)	24814	0.99 (0.95 to 1.03)
Chest pain										
Annual mean NO_2 (3.2 ppb)	28194	0.96 (0.88 to 1.05)	23315	0.96 (0.87 to 1.06)	20823	1.00 (0.87 to 1.15)	16279	0.90 (0.80 to 1.01)	17868	0.94 (0.84 to 1.06)
Distance to major road (doubling)	28194	0.98 (0.95 to 1.02)	23315	1.00 (0.96 to 1.04)	20823	1.00 (0.96 to 1.04)	16279	0.94 (0.98 to 1.10)	17868	1.02 (0.97 to 1.07)
Distance to minor road (doubling)	28194	0.98 (0.95 to 1.01)	23315	1.00 (0.96 to 1.03)	20823	0.99 (0.95 to 1.02)	16279	0.97 (0.90 to 1.05)	17868	0.99 (0.94 to 1.04)
Palpitations										
Annual mean NO_2 (3.7 ppb)	35809	0.80 (0.74 to 0.87)	31176	0.99 (0.91 to 1.08)	30412	1.00 (0.90 to 1.12)	20917	0.96 (0.87 to 1.07)	20935	0.96 (0.85 to 1.07)
Distance to major road (doubling)	35809	1.04 (1.02 to 1.06)	33062	1.00 (0.98 to 1.03)	30412	1.01 (0.98 to 1.03)	22167	1.01 (0.97 to 1.04)	21669	1.02 (0.98 to 1.05)
Distance to minor road (doubling)	35809	1.08 (1.05 to 1.10)	33062	1.00 (0.97 to 1.02)	30412	0.99 (0.97 to 1.02)	22167	1.00 (0.96 to 1.05)	21669	0.98 (0.94 to 1.02)

Sample size including repeated observations (N) and ORs with 99% CIs in parentheses; models are two-level models unless specified otherwise.
*Adjusted for known risk factors and identified confounders, specifically: models of allergies adjusted for cohort and smoking status and additionally residential remoteness in the NO_2 model; models of breathing difficulty symptoms adjusted for cohort and smoking status, and additionally highest qualification in the NO_2 model; models of chest pain symptoms adjusted for cohort, BMI group, smoking status and physical activity; models of palpitations symptoms adjusted for cohort, BMI group, smoking status, physical activity and additionally highest qualification in the NO_2 model.
†Adjusted for all available covariates.
BMI, body mass index.

symptoms in a Swedish study.[34] Inconsistencies in findings between these cross-sectional studies are thought to be due to differences in exposure measurement, misclassification, migration and heterogeneous assessment of confounders.[39]

Cardiovascular outcomes

The adverse effects of PM air pollution on cardiovascular health are established[44] and the consensus of published evidence suggests a causal association between air pollution and cardiovascular disease even at concentrations below existing air quality standards.[45 46] While long-term effects on mortality due to PM have been well documented, the evidence for effects on cardiovascular morbidity have been less consistent.[8 47] In contrast to our findings for heart disease, positive cross-sectional associations have been observed with long-term exposure to PM[48] and high traffic exposure.[49] Cross-sectional associations with NO_2 pollution are less understood, with some studies reporting positive associations[50] and others no association.[51] Experimental studies suggest low cardiovascular toxicity of NO_2 at ambient concentrations[52 53] and inconsistencies in findings between observational studies are thought to be due to exposure to co-pollutants or the combined effects of pollutant mixtures.[44]

Although traffic-related pollution has been associated with increases in blood pressure,[54 55] few studies of long-term exposure to ambient air pollution and hypertension prevalence have been conducted. The effect of air pollution on blood pressure is considered a potential mechanism that may explain the established association between ambient air pollution and cardiovascular disease.[54 55] Our results are consistent with several previous studies that observed no association between hypertension prevalence and exposure to NO_2 or PM pollution,[55 56] but are inconsistent with the positive association with PM exposure observed by Johnson and Parker,[48] and the inverse association with exposure to NO_X observed by Sørensen *et al*.[57] The differences in findings between studies are thought to be due to differential diagnostic criteria or misclassification of hypertension cases which are often undiagnosed[54 55]; however, the latter is unlikely to be relevant in Australia where blood pressure measurement is frequent and widespread.

Long-term air pollution exposure has been associated with stroke hospitalisations and mortality,[58] and there is growing evidence that it has an impact on the development of carotid arteriosclerosis which is a precursor of stroke.[59] However, limited evidence has been presented for an association with stroke prevalence. Consistent with our finding, Dong *et al*[14] and Forbes *et al*[51] observed no association with ambient NO_2 air pollution among Chinese or English women, respectively, whereas associations with PM pollution were observed by Dong *et al*, but not by Forbes *et al*. Brauer[59] suggests a limitation in studies that fail to separate ischaemic from haemorrhagic stroke, as air pollution is considered to variably affect the underlying pathophysiological mechanisms.

Limitations

Our results should be considered in the context of the following limitations. Although the ALSWH study and its self-report data have been extensively documented and validated, the study was not conceived as an air pollution study. Moreover, rural areas which typically have low air pollution levels are considerably over-represented in the ALSWH, and the use of subject-specific sampling weights was precluded by our choice of methodology. While our NO_2 data have been validated,[19] we were not able to assess the validity of our road proximity variables.

Several avenues of exposure misclassification may be present. First, as this is a cross-sectional study of disease prevalence rather than incidence, there is uncertainty in the degree to which exposure preceded the outcomes observed. However, high correlation was observed between annual average pollution levels in successive years during the study period. Second, we defined an arbitrary threshold for residential mobility between surveys of 5 km. Even with the analysis limited to movement below the threshold, exposures may be misclassified for some respondents. However, sensitivity analyses revealed no appreciable change in our estimates with thresholds of 1–10 km. Lastly, we did not account for occupational and transport-related exposures, nor indoor sources of NO_2 such as unflued gas stoves and heaters. We believe that each of these avenues of exposure misclassification could be non-differential and that the likely consequences are effects that are biased towards the null. A further source of potential bias towards the null may be the uncertainty inherent in exposure estimates from our land-use regression model.[60] In addition, prevalence studies may be affected by exposure-induced migration patterns, with susceptible individuals moving to lower exposure areas.[33 61]

As the study was based on self-reported assessment of health status and health-related variables, there may be misclassification in the outcome and covariate data. Undiagnosed and untreated cases may not have been reported, and there may have been misreporting of diagnoses or sensitive health-related characteristics such as weight, level of physical activity and alcohol use. However, validation studies of ALSWH self-report data suggest substantial agreement with hospital records for diabetes and to a reasonable degree, for heart disease and stroke.[62] The validity of self-reported height, weight and physical activity have also been assessed and confirmed by previous studies.[63 64] Although we were able to adjust for a wide array of potential confounders, residual confounding may nonetheless exist due to the coarse resolution of some covariates such as smoking. In addition, we were not able to assess the effects of traffic noise and therefore, cannot rule out residual confounding by exposure to noise.

As spatial autocorrelation was observed in the NO_2 exposure, we accounted for clustering by postcode. While this assumption allows for non-zero correlation between women living within the same postcode, it does not allow for correlation between spatially proximal women living in differing postcodes. An inability to completely account for spatial autocorrelation may be a source of bias in our study. However, there appeared to be minimal clustering by postcode based on the magnitude of the estimated within-postcode correlation.

Although we concentrated on exposure to ambient NO_2 as a marker for combustion-derived air pollution, individuals are exposed to a mixture of pollutants. However, we also analysed residential proximity to roads as a marker for the diverse mix of traffic-related ambient air pollution and our conclusions remain unchanged. Finally, as we studied the relationship between long-term exposure to ambient air pollution and chronic morbidity in women, our results cannot be extrapolated to the short-term effects of ambient air pollution or generalised to effects in men.

CONCLUSION

We observed no evidence of association between estimated long-term ambient NO_2 exposure and self-reported diseases (diabetes, heart disease, hypertension, stroke, asthma, COPD) and symptoms (allergies, breathing difficulties, chest pain, palpitations) in a cohort of 26 991 Australian women born in 1921–1926, 1946–1951 and 1973–1978. The observed results may be due to an inherently low signal-to-noise ratio with uncertainty in both outcome and exposure measurement, and therefore the potential for misclassification which may have biased the results towards the null; lack of power to detect modest NO_2 effect sizes for some health conditions; or may reflect an absence of effects at the relatively low annual average NO_2 levels observed in Australia. Nonetheless, several of our findings are consistent with those observed in other work. Our study adds to the limited evidence base on the long-term effects of ambient NO_2 air pollution and traffic exposure on chronic cardiorespiratory, metabolic and allergic conditions in women.

Twitter Follow Adrian Barnett at @aidybarnett

Acknowledgements The research on which this paper is based was conducted as part of the Australian Longitudinal Study on Women's Health by the University of Newcastle and the University of Queensland. The authors are grateful to the women who provided their time and personal data in the surveys, and the Australian Government Department of Health for funding the ALSWH. They thank the ALSWH team for their assistance with the data.

Contributors NL performed the statistical analysis and prepared the manuscript. LDK coordinated the study and performed the exposure assessment. AJD and AGB guided the statistical analyses. LDK and AGB conceived the idea and initiated the study. All authors were involved in the study design, critically reviewed drafts, and have read and approved the final manuscript.

Funding LDK was supported by an NHMRC Early Career (Australian Public Health) Fellowship (APP1036620). The Australian Longitudinal Study on Women's Health is funded by the Australian Government Department of Health.

Competing interests None declared.

Ethics approval The ALSWH is approved by the human research ethics committees of the University of Queensland and the University of Newcastle.

Provenance and peer review Not commissioned; externally peer reviewed.

Data sharing statement ALSWH and NO_2 exposure data are available (see http://www.alswh.org.au and Knibbs *et al*[19]).

REFERENCES

1. Lim SS, Vos T, Flaxman AD, *et al*. A comparative risk assessment of burden of disease and injury attributable to 67 risk factors and risk factor clusters in 21 regions, 1990–2010: a systematic analysis for the Global Burden of Disease Study 2010. *Lancet* 2012;380:2224–60.
2. Cohen AJ, Anderson HR, Ostro B, *et al*. The global burden of disease due to outdoor air pollution. *J Toxicol Environ Health A* 2005;68:1301–7.
3. Hansen A, Bi P, Nitschke M. Air pollution and cardiorespiratory health in Australia: the impact of climate change. *Environ Health* 2009;9:17.
4. Balti EV, Echouffo-Tcheugui JB, Yako YY, *et al*. Air pollution and risk of type 2 diabetes mellitus: a systematic review and meta-analysis. *Diabetes Res Clin Pract* 2014;106:161–72.
5. Anderson HR, Favarato G, Atkinson RW. Long-term exposure to air pollution and the incidence of asthma: meta-analysis of cohort studies. *Air Qual Atmos Health* 2011;6:47–56.
6. Mustafic H, Jabre P, Caussin C, *et al*. Main air pollutants and myocardial infarction: a systematic review and meta-analysis. *JAMA* 2012;307:713–21.
7. Yang W-S, Wang X, Deng Q, *et al*. An evidence-based appraisal of global association between air pollution and risk of stroke. *Int J Cardiol* 2014;175:307–13.
8. Brook RDR, Rajagopalan SS, Pope CAC, *et al*. Particulate matter air pollution and cardiovascular disease: an update to the scientific statement from the American Heart Association. *Circulation* 2010;121:2331–78.
9. Gowers AM, Cullinan P, Ayres JG, *et al*. Does outdoor air pollution induce new cases of asthma? Biological plausibility and evidence; a review. *Respirology* 2012;17:887–98.
10. Carlsten C, Melén E. Air pollution, genetics, and allergy: an update. *Curr Opin Allergy Clin Immunol* 2012;12:455–60.
11. Brook RD, Jerrett M, Brook JR, *et al*. The relationship between diabetes mellitus and traffic-related air pollution. *J Occup Environ Med* 2008;50:32–8.
12. Chen LH, Knutsen SF, Shavlik D, *et al*. The association between fatal coronary heart disease and ambient particulate air pollution: are females at greater risk? *Environ Health Perspect* 2005;113:1723–9.
13. Künzli N, Jerrett M, Mack WJ, *et al*. Ambient air pollution and atherosclerosis in Los Angeles. *Environ Health Perspect* 2005;113:201–6.
14. Dong G-H, Qian ZM, Wang J, *et al*. Associations between ambient air pollution and prevalence of stroke and cardiovascular diseases in 33 Chinese communities. *Atmospheric Environ* 2013;77:968–73.
15. Lee C, Dobson AJ, Brown WJ, *et al*. Cohort profile: the Australian Longitudinal Study on Women's Health. *Int J Epidemiol* 2005;34:987–91.
16. Lipsitch M, Tchetgen Tchetgen E, Cohen T. Negative controls: a tool for detecting confounding and bias in observational studies. *Epidemiology* 2010;21:383–8.
17. Lindeboom M, van Doorslaer E. Cut-point shift and index shift in self-reported health. *J Health Econ* 2004;23:1083–99.
18. Groot W. Adaptation and scale of reference bias in self-assessments of quality of life. *J Health Econ* 2000;19:403–20.
19. Knibbs LD, Hewson MG, Bechle MJ, *et al*. A national satellite-based land-use regression model for air pollution exposure assessment in Australia. *Environ Res* 2014;135:204–11.

20. Jerrett M, Arain A, Kanaroglou P, *et al*. A review and evaluation of intraurban air pollution exposure models. *J Expo Anal Environ Epidemiol* 2005;15:185–204.
21. Zou G. A modified Poisson regression approach to prospective studies with binary data. *Am J Epidemiol* 2004;159:702–6.
22. Yelland LN, Salter AB, Ryan P. Performance of the modified Poisson regression approach for estimating relative risks from clustered prospective data. *Am J Epidemiol* 2011;174:984–92.
23. Australian Institute of Health and Welfare. *Risk factors contributing to chronic disease*. 2012 (cited 27 April 2015); AIHW cat. no. PHE157. Canberra: AIHW. http://www.aihw.gov.au/publication-detail/?id=10737421466
24. Rothman KJ, Greenland S, Lash TL. *Modern epidemiology*. Philadelphia: Lippincott Williams & Wilkins, 2008. http://www.lww.com/Product/9781451190052
25. Lee PH. Is a cutoff of 10% appropriate for the change-in-estimate criterion of confounder identification? *J Epidemiol* 2014;24:161–7.
26. Pan W. Akaike's information criterion in generalized estimating equations. *Biometrics* 2001;57:120–5.
27. Cui J. QIC program and model selection in GEE analyses. *Stata J* 2007;7:209–20.
28. Eze IC, Schaffner E, Fischer E, *et al*. Long-term air pollution exposure and diabetes in a population-based Swiss cohort. *Environ Int* 2014;70:95–105.
29. Janghorbani M, Momeni F, Mansourian M. Systematic review and metaanalysis of air pollution exposure and risk of diabetes. *Eur J Epidemiol* 2014;29:231–42.
30. Dijkema MBA, Mallant SF, Gehring U, *et al*. Long-term exposure to traffic-related air pollution and type 2 diabetes prevalence in a cross-sectional screening-study in the Netherlands. *Environ Health* 2011;10:76.
31. Kim K-H, Jahan SA, Kabir E. A review on human health perspective of air pollution with respect to allergies and asthma. *Environ Int* 2013;59:41–52.
32. Guarnieri M, Balmes JR. Outdoor air pollution and asthma. *Lancet* 2014;383:1581–92.
33. Anderson HR, Favarato G, Atkinson RW. Long-term exposure to outdoor air pollution and the prevalence of asthma: meta-analysis of multi-community prevalence studies. *Air Qual Atmos Health* 2011;6:57–68.
34. Lindgren A, Stroh E, Montnémery P, *et al*. Traffic-related air pollution associated with prevalence of asthma and COPD/chronic bronchitis. A cross-sectional study in Southern Sweden. *Int J Health Geogr* 2009;8:2–2.
35. Lindgren A, Stroh E, Nihlén U, *et al*. Traffic exposure associated with allergic asthma and allergic rhinitis in adults. A cross-sectional study in southern Sweden. *Int J Health Geogr* 2009;8:25.
36. Cesaroni G, Badaloni C, Porta D, *et al*. Comparison between various indices of exposure to traffic-related air pollution and their impact on respiratory health in adults. *Occup Environ Med* 2008;65:683–90.
37. Pujades-Rodríguez M, Lewis S, Mckeever T, *et al*. Effect of living close to a main road on asthma, allergy, lung function and chronic obstructive pulmonary disease. *Occup Environ Med* 2009;66:679–84.
38. Bui DS, Burgess JA, Matheson MC, *et al*. Ambient wood smoke, traffic pollution and adult asthma prevalence and severity. *Respirology* 2013;18:1101–7.
39. Schikowski T, Mills IC, Anderson HR, *et al*. Ambient air pollution: a cause of COPD? *Eur Respir J* 2014;43:250–63.
40. Götschi T, Heinrich J, Sunyer J, *et al*. Long-term effects of ambient air pollution on lung function: a review. *Epidemiology* 2008;19:690–701.
41. Cai Y, Schikowski T, Adam M, *et al*. Cross-sectional associations between air pollution and chronic bronchitis: an ESCAPE meta-analysis across five cohorts. *Thorax* 2014;69:1005–14.
42. Song Q, Christiani DC, Wang XR, *et al* The global contribution of outdoor air pollution to the incidence, prevalence, mortality and hospital admission for chronic obstructive pulmonary disease: a systematic review and meta-analysis. *Int J Environ Res Public Health* 2014;11:11822–32.
43. Schikowski T, Sugiri D, Ranft U, *et al*. Long-term air pollution exposure and living close to busy roads are associated with COPD in women. *Respir Res* 2005;6:152.
44. Shah AS, Langrish JP, Nair H, *et al*. Global association of air pollution and heart failure: a systematic review and meta-analysis. *Lancet* 2013;382:1039–48.
45. Simkhovich BZ, Kleinman MT, Kloner RA. Air pollution and cardiovascular injury epidemiology, toxicology, and mechanisms. *J Am Coll Cardiol* 2008;52:719–26.
46. Mills NL, Donaldson K, Hadoke PW, *et al*. Adverse cardiovascular effects of air pollution. *Nat Clin Pract Cardiovasc Med* 2009;6:36–44.
47. Bhaskaran K, Hajat S, Haines A, *et al*. The effects of air pollution on the incidence of myocardial infarction. *Heart* 2009;95:1746–59.
48. Johnson D, Parker JD. Air pollution exposure and self-reported cardiovascular disease. *Environ Res* 2009;109:582–9.
49. Hoffmann B, Moebus S, Stang A, *et al*. Residence close to high traffic and prevalence of coronary heart disease. *Eur Heart J* 2006;27:2696–702.
50. Beckerman BS, Jerrett M, Finkelstein M, *et al*. The association between chronic exposure to traffic-related air pollution and ischemic heart disease. *J Toxicol Environ Health A* 2012;75:402–11.
51. Forbes LJL, Patel MD, Rudnicka AR, *et al*. Chronic exposure to outdoor air pollution and diagnosed cardiovascular disease: meta-analysis of three large cross-sectional surveys. *Environ Health* 2009;8:30.
52. Langrish JP, Lundbäck M, Barath S, *et al*. Exposure to nitrogen dioxide is not associated with vascular dysfunction in man. *Inhal Toxicol* 2010;22:192–8.
53. Mills NL, Miller MR, Lucking AJ, *et al*. Combustion-derived nanoparticulate induces the adverse vascular effects of diesel exhaust inhalation. *Eur Heart J* 2011;32:2660–71.
54. Chan SH, Van Hee VC, Bergen S, *et al*. Long-term air pollution exposure and blood pressure in the Sister Study. *Environ Health Perspect* 2015;123:951–8.
55. Foraster M, Basagaña X, Aguilera I, *et al*. Association of long-term exposure to traffic-related air pollution with blood pressure and hypertension in an adult population-based cohort in Spain (the REGICOR study). *Environ Health Perspect* 2014;122:404–11.
56. Dong G-H, Qian ZM, Xaverius PK, *et al*. Association between long-term air pollution and increased blood pressure and hypertension in China. *Hypertension* 2013;61:578–84.
57. Sørensen M, Hoffmann B, Hvidberg M, *et al*. Long-term exposure to traffic-related air pollution associated with blood pressure and self-reported hypertension in a Danish cohort. *Environ Health Perspect* 2012;120:418–24.
58. Ljungman PL, Mittleman MA. Ambient air pollution and stroke. *Stroke* 2014;45:3734–41.
59. Brauer M. Air pollution, stroke, and anxiety. *BMJ* 2015;350:h1510–0.
60. Basagaña X, Aguilera I, Rivera M, *et al*. Measurement error in epidemiologic studies of air pollution based on land-use regression models. *Am J Epidemiol* 2013;178:1342–6.
61. Jacquemin B, Schikowski T, Carsin AE, *et al*. The role of air pollution in adult-onset asthma: a review of the current evidence. *Semin Respir Crit Care Med* 2012;33:606–19.
62. Navin Cristina TJ, Stewart Williams JA, Parkinson L, *et al*. Identification of diabetes, heart disease, hypertension and stroke in mid- and older-aged women: comparing self-report and administrative hospital data records. *Geriatr Gerontol Int* 2015.
63. Brown WJ, Burton NW, Marshall AL, *et al*. Reliability and validity of a modified self-administered version of the Active Australia physical activity survey in a sample of mid-age women. *Aust N Z J Public Health* 2008;32:535–41.
64. Burton NW, Brown W, Dobson A. Accuracy of body mass index estimated from self-reported height and weight in mid-aged Australian women. *Aust N Z J Public Health* 2010;34:620–3.

Supplementary material: Long-term ambient air pollution exposure and self-reported morbidity in the Australian Longitudinal Study on Women's Health: a cross-sectional study

Nina Lazarevic, Annette J. Dobson, Adrian G. Barnett, Luke D. Knibbs

1 Definition of outcome measures

There were some differences between the exact questions asked across surveys which introduced some ambiguity in the definition of the pooled outcome variables. While only type 2 diabetes cases were included in the younger cohort, type 1 and type 2 diabetes were not differentiated in surveys of the middle-aged and older cohorts, however the majority of cases in the older age groups were assumed to be type 2 diabetes.

In surveys of the middle-aged and older cohorts, the COPD question referred to diagnosis or treatment of bronchitis or emphysema, whereas in surveys of the younger cohort the presence of emphysema was not assessed. Data were nevertheless pooled across the three age groups as rates of emphysema in the younger cohort were expected to be negligible.

2 Residential mobility threshold sensitivity analysis

For the disease prevalence models, a sensitivity analysis was performed to determine whether the results were affected by the choice of an arbitrary 5 km threshold on residential mobility. Models adjusted for known risk factors and confounders were re-estimated with varying thresholds between 1 km and 10 km, in 0.5 km increments.

3 Covariates assessed for confounding

Covariates assessed for confounding of the association between outcome and the exposure variables of interest included:

- age group (cohort membership);
- body mass index, defined as weight in kilograms divided by squared height in metres, and categorised according to the World Health Organisation classification[1] (underweight <18.5 kg/m^2, healthy weight ≥ 18.5 and <25 kg/m^2, overweight ≥ 25 and <30 kg/m^2, and obese ≥ 30 kg/m^2);

[1] World Health Organization. Obesity: Preventing and Managing the Global Epidemic[Internet]. 2000 [cited 2015 Apr 27]; Geneva: WHO. Available from: http://www.who.int/nutrition/publications/obesity/WHO_TRS_894/en/

- smoking status, categorised as never-smoker, ex-smoker, or current smoker;
- alcohol use, with categories based on the Australian National Health and Medical Research Council classification[2,3]: non-drinker, rarely drinks ($<$2 standard drinks per fortnight), low-risk drinker (up to 14 standard drinks per week), risky or high risk drinker ($\geq$15 standard drinks per week);
- level of physical activity, categorised[4] as nil/sedentary ($<$10 minutes of moderate activity per week), low (11-150 minutes of moderate activity per week), moderate (151-300 minutes of moderate activity per week), or high ($>$300 minutes of moderate activity per week);
- marital status, categorised as (1) married or in a de-facto relationship, (2) separated, widowed or divorced, and (3) single;
- vegetable consumption, categorised as 1 serve or less, 2-3 serves, 4 serves, or 5 or more serves of vegetables per day;
- fruit consumption, categorised as 1 serve or less, 2-3 serves, or 4 or more serves of fruit per day;
- degree of urbanisation or remoteness, with categories based on the Accessibility Remoteness Index for Australia (ARIA+) classification[5], (1) major cities, (2) inner regional areas, (3) outer regional areas, or (4) remote or very remote areas;
- mean annual temperature at the place of residence;
- highest qualification achieved, a proxy variable for socioeconomic status, categorised as (1) no formal education, (2) high school certificate or less, (3) trade, apprenticeship, certificate or diploma, and (4) university degree or higher; and
- the self-assessed ability to manage on income available, another proxy variable for socioeconomic status, categorised on a 5-point ordinal scale as (1) impossible, (2) difficult always, (3) difficult sometimes, (4) not too bad, (5) easy.

BMI was analysed as a categorical rather than continuous variable because it was expected to have a non-linear relationship with some of the outcome variables considered.

4 Three-level mixed model of symptom prevalence

Subjects were considered nested within postcodes and the analysis was restricted to respondents below the 5 km threshold for residential mobility, with the postcode at the first survey retained for the analysis. Where the postcode in the first of the two surveys was missing, the postcode in the second survey was assigned to the respondent, provided the distance moved was less than the chosen 5 km threshold. Estimating a crossed effects model was attempted to avoid the necessity of choosing a postcode for mobile respondents, but not successful due to numerical difficulties. As the analysis which attempted to account

[2] National Health and Medical Research Council. Australian Alcohol Guidelines: Health Risks and Benefits[Internet]. 2001 [cited 2015 Apr 27]; Canberra: Commonwealth of Australia. Available from: https://www.nhmrc.gov.au/guidelines-publications/ds9

[3] ALSWH Data Dictionary Supplement: Alcohol intake[Internet]. [cited 2015 Apr 27] Available from: http://www.alswh.org.au/images/content/pdf/InfoData/Data_Dictionary_Supplement/DDSSection2AlcIntake.pdf

[4] ALSWH Data Dictionary Supplement: Physical activity[Internet]. [cited 2015 Apr 27] Available from: http://www.alswh.org.au/images/content/pdf/InfoData/Data_Dictionary_Supplement/DDSSection2PhysicalActivityS2.pdf

[5] Australian Institute of Health and Welfare. Rural, regional and remote health: a guide to remoteness classifications[Internet]. 2004 [cited 2015 Apr 27]; AIHW cat. no. PHE 53. Canberra: AIHW. Available from: http://www.aihw.gov.au/publication-detail/?id=6442467589

for spatial autocorrelation required a substantial loss in sample size, the two-level model was chosen for the purposes of model building and only the final model was estimated with the additional random intercept, for comparison. Parameter estimates from the two-level model were used as starting values in the optimisation algorithm.

5 Goodness-of-fit

In the disease prevalence models, goodness-of-fit was assessed by the Wald χ^2 statistic and a modified Hosmer-Lemeshow test statistic for population-averaged GEE models, using 10 groups[6]. A Bonferroni correction was applied at a 5% significance level to account for the 9 simultaneous tests. Tests of adjusted models revealed some possible lack of fit in the following models: diabetes risk and proximity to major and minor road models (p=0.003); heart disease risk and NO_2 (p=0.002); and all hypertension models (p<0.001).

Dispersion was assessed by dividing the Pearson χ^2 statistic by the model degrees of freedom. All Pearson dispersion statistics were between 0.9 and 0.94 except in models of hypertension prevalence, which were moderately under-dispersed with statistics of 0.68.

6 Supplementary Tables and Figures

Table S1 shows the derivation of analytical sample sizes with number of missing observations by outcome.

Tables S2 and S3 show descriptive statistics by outcome for the self-reported disease and symptom data, respectively.

Tables S4 and S5 show results from the power analysis for the NO_2 exposure and proximity to roads exposures, respectively.

Figures S1 and S2 show the results for the residential mobility threshold sensitivity analysis.

[6]Horton NJ, Bebchuk JD, Jones CL, et al. Goodness-of-fit for GEE: an example with mental health service utilization. Statist Med 1999;18:213-22. doi: 10.1002/(SICI)1097-0258(19990130)18:2<213::AID-SIM999>3.0.CO;2-E

Table S1: Derivation of analytical sample size and number of missing observations.

	Diabetes[1]	Heart disease[1]	Hypertension[1]	Stroke[2]	Asthma[1]	COPD[1]	Allergies[3]	Breathing difficulties[4]	Chest pain[5]	Palpitations[3]
Available sample size	22266	22266	22266	14066	22266	22266	37994	38464	30264	37994
Outcome										
Not missing	21482	21482	21482	13913	21482	21482	37493	37250	28974	37497
Missing	784	784	784	153	784	784	501	1214	1290	497
Exposure† *(NO_2)*										
Not missing	19777	19777	19777	13151	19777	19777	35797	35457	28194	35809
Missing	1705	1705	1705	762	1705	1705	1696	1793	780	1688
Postcode (alias)†										
Not missing	19536	19536	19536	13001	19536	19536	35523	35115	27955	35538
Missing	241	241	241	150	241	241	274	342	239	271
Residential mobility†										
Move $\leq$ 5 km	14563	14563	14563	10402	14563	14563	23286	24818	20241	23299
Move $>$ 5 km	4973	4973	4973	2599	4973	4973	12237	10297	7714	12239
Analytical sample size‡										
2-level model	14563	14563	14563	10402	14563	14563	35797	35457	28194	35809
3-level model							23286	24818	20241	23299
Covariates - number missing§										
Marital status	83	83	83	65	83	83	207	205	176	206
BMI group	848	848	848	729	848	848	1059	2174	1949	1068
Smoking status	74	74	74	61	74	74	121	1839	1801	118
Vegetable serves per day	68	68	68	68	68	68	109	177	174	111
Fruit serves per day	51	51	51	51	51	51	84	137	136	80
Physical activity	828	828	828	709	828	828	1707	2117	1881	1716
Highest qualification	584	584	584	499	584	584	2221	2513	2357	2222
Alcohol use	97	97	97	89	97	97	237	1935	1882	235
Remoteness (ARIA+ group)	9	9	9	1	9	9	180	300	140	180
Ability to manage on income	103	103	103	81	103	103	199	241	207	200
Annual mean temperature (°C)	69	69	69	54	69	69	167	164	129	165

The upper section of the table is negatively cumulative in the sense that it shows for each outcome, the number of non-missing observations by exposure; then of those, the number of non-missing observations by postcode; and then of those, the number above and below the chosen threshold for residential mobility. Figures for symptom data include repeated observations.

† Of the non-missing observations in the variable shown immediately above in the table.

‡ For the disease data (2-level model) and symptom data (3-level model): sample size with non-missing outcome, exposure, and postcode variables; and with a 5 km threshold on residential mobility between surveys. For the symptom data (2-level model): sample size with non-missing outcome and exposure.

§ Of the analytical sample size for the 2-level model.

1 Data are survey 5 of the younger cohort and surveys 6 of the middle-aged and older cohorts.

2 Data are surveys 6 of the middle-aged and older cohorts.

3 Data are survey 4 and survey 5 of the younger cohort and survey 5 and survey 6 of the middle-aged cohort.

4 Data are survey 5 of the younger cohort and surveys 5 and surveys 6 of the middle-aged and older cohorts.

5 Data are survey 5 and surveys 6 of the middle-aged and older cohorts.

Table S2: Descriptive statistics of the self-reported disease data.

	Diabetes[†]		Heart disease[†]		Hypertension[†]		Stroke[‡]		Asthma[†]		COPD[†]	
	Prevalence	N[§]	Prevalence	N[§]	Prevalence	N[§]	Prevalence	N[§]	Prevalence	N[§]	Prevalence	N[§]
Overall	0.069	14563	0.086	14563	0.294	14563	0.019	10402	0.106	14563	0.063	14563
Cohort												
Younger	0.012	4161	0.003	4161	0.050	4161			0.108	4161	0.064	4161
Middle-aged	0.079	7223	0.045	7223	0.301	7223	0.009	7223	0.114	7223	0.067	7223
Older	0.122	3179	0.289	3179	0.599	3179	0.043	3179	0.085	3179	0.054	3179
Smoking status												
Never smoker	0.072	9243	0.090	9243	0.307	9243	0.020	6712	0.100	9243	0.053	9243
Ex-smoker	0.066	3961	0.092	3961	0.299	3961	0.020	2956	0.110	3961	0.070	3961
Current smoker	0.061	1285	0.039	1285	0.183	1285	0.012	673	0.137	1285	0.113	1285
BMI group (WHO)												
Underweight	0.052	388	0.139	388	0.265	388	0.062	276	0.103	388	0.103	388
Healthy weight	0.037	5958	0.077	5958	0.216	5958	0.018	3886	0.083	5958	0.052	5958
Overweight	0.061	4260	0.081	4260	0.299	4260	0.016	3230	0.107	4260	0.056	4260
Obese	0.139	3109	0.081	3109	0.406	3109	0.017	2281	0.156	3109	0.091	3109
Physical activity												
Nil/sedentary	0.114	3288	0.190	3288	0.450	3288	0.038	2762	0.118	3288	0.074	3288
Low	0.062	4095	0.063	4095	0.272	4095	0.016	2486	0.103	4095	0.063	4095
Moderate	0.051	2651	0.051	2651	0.227	2651	0.008	1753	0.097	2651	0.055	2651
High	0.047	3701	0.039	3701	0.220	3701	0.010	2692	0.101	3701	0.055	3701
Marital status												
Married/De facto	0.056	9494	0.047	9494	0.236	9494	0.012	6235	0.105	9494	0.061	9494
Separated/Divorced/Widowed	0.110	4014	0.195	4014	0.471	4014	0.033	3821	0.106	4014	0.065	4014
Single	0.033	972	0.027	972	0.133	972	0.007	281	0.114	972	0.071	972
Alcohol use												
Non-drinker	0.122	3105	0.166	3105	0.414	3105	0.038	2555	0.108	3105	0.064	3105
Rarely drinks	0.085	3274	0.081	3274	0.297	3274	0.016	2234	0.123	3274	0.072	3274
Low risk drinker	0.044	7410	0.059	7410	0.241	7410	0.012	5024	0.098	7410	0.057	7410
Risky or high risk drinker	0.025	677	0.046	677	0.300	677	0.014	500	0.102	677	0.080	677
Vegetables serves per day												
1 serve or less	0.086	1979	0.128	1979	0.360	1979	0.028	1804	0.113	1979	0.061	1979
2-3 serves	0.072	7704	0.081	7704	0.291	7704	0.019	5556	0.105	7704	0.067	7704
4 serves	0.062	3020	0.084	3020	0.295	3020	0.015	2003	0.103	3020	0.055	3020
5 or more serves	0.053	1792	0.062	1792	0.228	1792	0.013	971	0.107	1792	0.062	1792

(Continues on next page.)

Descriptive statistics of the self-reported disease data. (continued)

	Diabetes†		Heart disease†		Hypertension†		Stroke‡		Asthma†		COPD†	
	Prevalence	N§	Prevalence	N§	Prevalence	N§	Prevalence	N§	Prevalence	N§	Prevalence	N§
Fruit serves per day												
1 serve or less	0.056	6074	0.073	6074	0.262	6074	0.025	3614	0.114	6074	0.070	6074
2-3 serves	0.078	7430	0.094	7430	0.312	7430	0.017	5812	0.100	7430	0.059	7430
4 or more serves	0.078	1008	0.108	1008	0.347	1008	0.011	925	0.102	1008	0.048	1008
Highest qualification												
No formal qualification	0.140	1846	0.178	1846	0.460	1846	0.032	1807	0.113	1846	0.065	1846
Year 12 or less	0.087	5541	0.108	5541	0.359	5541	0.018	4665	0.102	5541	0.064	5541
Trade/App./Cert./Dip.	0.045	2970	0.061	2970	0.239	2970	0.018	1849	0.115	2970	0.069	2970
University degree or higher	0.023	3622	0.026	3622	0.148	3622	0.010	1582	0.094	3622	0.053	3622
Ability to manage on income												
Impossible	0.112	224	0.080	224	0.348	224	0.041	170	0.196	224	0.098	224
Difficult always	0.100	1245	0.075	1245	0.296	1245	0.019	832	0.155	1245	0.108	1245
Difficult sometimes	0.080	3219	0.065	3219	0.255	3219	0.015	2005	0.130	3219	0.081	3219
Not too bad	0.064	6418	0.089	6418	0.303	6418	0.020	4792	0.097	6418	0.054	6418
Easy	0.052	3354	0.106	3354	0.310	3354	0.021	2522	0.076	3354	0.045	3354
Remoteness (ARIA+ group)												
Major cities	0.057	7168	0.082	7168	0.262	7168	0.021	4589	0.103	7168	0.066	7168
Inner regional	0.080	5022	0.090	5022	0.323	5022	0.019	3965	0.114	5022	0.063	5022
Outer regional	0.079	2068	0.093	2068	0.331	2068	0.014	1617	0.093	2068	0.052	2068
Remote or very remote	0.115	296	0.095	296	0.351	296	0.030	230	0.118	296	0.054	296
	Cases	Non-cases	Cases	Non-cases	Cases	Non-cases	Cases	Non-cases	Cases	Non-cases	Cases	Non-cases
Sample size§	1011	13552	1258	13305	4283	10280	202	10200	1543	13020	918	13645
Annual mean temperature (°C)	16.4 (12.0-22.0)	16.4 (12.0-21.0)	16.3 (12.0-21.0)	16.4 (12.0-22.0)	16.3 (12.0-21.0)	16.4 (12.0-22.0)	16.3 (12.0-20.0)	16.4 (12.0-21.0)	16.4 (12.0-21.0)	16.4 (12.0-22.0)	16.4 (12.0-21.0)	16.4 (12.0-22.0)
3-year mean NO_2 (ppb)	5.2 (2.4-9.8)	5.7 (2.4-11.3)	5.3 (2.4-10.0)	5.7 (2.4-11.4)	5.2 (2.4-10.2)	5.8 (2.4-11.6)	5.2 (2.6-9.5)	5.2 (2.3-10.3)	5.5 (2.5-10.9)	5.6 (2.4-11.3)	5.7 (2.5-10.7)	5.6 (2.4-11.3)
Distance to major road (km)	1.0 (0.0-1.7)	0.5 (0.0-1.6)	0.5 (0.0-1.4)	0.6 (0.0-1.6)	0.6 (0.0-1.6)	0.6 (0.0-1.6)	0.6 (0.0-1.2)	0.6 (0.0-1.7)	0.6 (0.0-1.7)	0.6 (0.0-1.6)	0.5 (0.0-1.5)	0.6 (0.0-1.6)
Distance to minor road (km)	0.1 (0.0-0.2)	0.1 (0.0-0.2)	0.1 (0.0-0.2)	0.1 (0.0-0.2)	0.1 (0.0-0.2)	0.1 (0.0-0.2)	0.1 (0.0-0.2)	0.1 (0.0-0.3)	0.1 (0.0-0.2)	0.1 (0.0-0.2)	0.1 (0.0-0.2)	0.1 (0.0-0.2)

For categorical variables, sample prevalence and sample size are shown; for continuous variables, mean with 5th to 95th percentile range in parentheses for cases and non-cases.

† Data are survey 5 of the younger cohort and surveys 6 of the middle-aged and older cohorts.

‡ Data are surveys 6 of the middle-aged and older cohorts.

§ Sample size with non-missing outcome, exposure, and alias postcode variables; and assuming a 5 km threshold on residential mobility between surveys.

Table S3: Descriptive statistics of the self-reported symptom data.

	Allergies†		Breathing difficulties‡		Chest pain#		Palpitations†	
	Prevalence	N§	Prevalence	N§	Prevalence	N§	Prevalence	N§
Overall	0.420	35797	0.161	35457	0.100	28194	0.161	35809
Cohort								
Younger	0.435	16060	0.078	7036			0.114	15980
Middle-aged	0.408	19737	0.153	19531	0.093	19375	0.199	19829
Older			0.245	8890	0.115	8819		
Smoking status								
Never smoker	0.414	21471	0.132	21253	0.090	16880	0.147	21484
Ex-smoker	0.442	9524	0.179	9324	0.104	7491	0.178	9509
Current smoker	0.405	4681	0.216	3041	0.121	2022	0.191	4698
BMI group (WHO)								
Underweight	0.387	746	0.213	879	0.108	683	0.199	742
Healthy weight	0.407	15447	0.124	14352	0.088	10694	0.147	15446
Overweight	0.425	10317	0.158	10447	0.098	8682	0.162	10339
Obese	0.442	8228	0.214	7605	0.119	6186	0.183	8214
Physical activity								
Nil/sedentary	0.417	4902	0.259	8251	0.139	7240	0.218	4902
Low	0.430	11081	0.140	9600	0.095	6845	0.155	11071
Moderate	0.425	7511	0.122	6347	0.081	4829	0.146	7527
High	0.410	10596	0.116	9142	0.075	7399	0.148	10593
Marital status								
Married/De facto	0.417	27273	0.133	22931	0.090	17357	0.156	27288
Separated/Divorced/Widowed	0.421	4565	0.228	10390	0.116	9915	0.217	4569
Single	0.446	3752	0.137	1931	0.101	746	0.124	3746
Alcohol consumption								
Non-drinker	0.407	4585	0.198	6997	0.118	6099	0.180	4587
Rarely drinks	0.423	8535	0.171	7433	0.106	5655	0.174	8541
Low risk drinker	0.420	20543	0.125	17432	0.083	13198	0.147	20549
Risky or high risk drinker	0.431	1897	0.169	1660	0.092	1360	0.209	1897
Vegetables serves per day								
1 serve or less	0.417	6359	0.227	5505	0.121	5163	0.193	6370
2-3 serves	0.424	18801	0.157	18203	0.098	14553	0.155	18803
4 serves	0.414	6661	0.139	7441	0.091	5601	0.150	6659
5 or more serves	0.421	3867	0.132	4131	0.089	2703	0.155	3866

(Continues on next page.)

Descriptive statistics of the self-reported symptom data. (continued)

	Allergies†		Breathing difficulties‡		Chest pain#		Palpitations†	
	Prevalence	N§	Prevalence	N§	Prevalence	N§	Prevalence	N§
Fruit serves per day								
1 serve or less	0.438	16302	0.179	13907	0.116	9677	0.166	16297
2-3 serves	0.409	17255	0.150	18649	0.093	15775	0.156	17265
4 or more serves	0.374	2156	0.143	2764	0.082	2606	0.164	2167
Highest qualification								
No formal qualification	0.402	2595	0.246	4948	0.137	4821	0.264	2622
Year 12 or less	0.415	11463	0.173	13695	0.096	12192	0.182	11480
Trade/App./Cert./Dip.	0.441	8079	0.143	6672	0.092	4822	0.163	8083
University degree or higher	0.417	11439	0.091	7629	0.066	4002	0.103	11402
Ability to manage on income								
Impossible	0.458	602	0.297	545	0.183	443	0.266	606
Difficult always	0.466	3785	0.244	3056	0.192	2318	0.252	3773
Difficult sometimes	0.446	9252	0.186	7663	0.118	5604	0.176	9240
Not too bad	0.407	14939	0.148	15670	0.088	12837	0.146	14968
Easy	0.386	7020	0.123	8282	0.069	6785	0.117	7022
Remoteness (ARIA+ group)								
Major cities	0.424	16490	0.157	15268	0.095	11255	0.147	16469
Inner regional	0.429	11979	0.168	12757	0.101	10854	0.177	12009
Outer regional	0.394	5950	0.162	6080	0.102	5113	0.170	5956
Remote or very remote	0.429	1198	0.130	1052	0.115	832	0.154	1195
	Cases	Non-cases	Cases	Non-cases	Cases	Non-cases	Cases	Non-cases
Sample size§	21330	14467	10346	25111	6452	21742	12303	23506
Annual mean temperature (°C)	16.5 (12.0-22.0)	16.5 (12.0-22.0)	16.4 (12.0-22.0)	16.4 (12.0-22.0)	16.5 (12.0-22.0)	16.4 (12.0-21.0)	16.5 (12.0-22.0)	16.5 (12.0-22.0)
Annual mean NO_2 (ppb)	5.8 (2.5-11.8)	5.8 (2.5-12.0)	5.5 (2.5-10.6)	5.5 (2.4-11.2)	5.2 (2.4-10.2)	5.3 (2.4-10.6)	5.5 (2.5-11.0)	5.9 (2.5-12.1)
Distance to major road (km)	0.7 (0.0-2.0)	0.8 (0.0-2.0)	0.7 (0.0-2.0)	0.7 (0.0-2.0)	0.8 (0.0-2.0)	0.7 (0.0-2.1)	0.7 (0.0-2.0)	0.7 (0.0-2.0)
Distance to minor road (km)	0.1 (0.0-0.4)	0.1 (0.0-0.4)	0.1 (0.0-0.4)	0.1 (0.0-0.5)	0.1 (0.0-0.5)	0.1 (0.0-0.6)	0.1 (0.0-0.4)	0.1 (0.0-0.4)

For categorical variables, sample prevalence and sample size are shown, including repeated observations; for continuous variables, mean with 5th to 95th percentile range in parentheses for cases and non-cases.

† Data are survey 4 and survey 5 of the younger cohort and survey 5 and survey 6 of the middle-aged cohort.

‡ Data are survey 5 of the younger cohort and surveys 5 and surveys 6 of the middle-aged and older cohorts.

Data are survey 5 and surveys 6 of the middle-aged and older cohorts.

§ Sample size, including repeated observations, with non-missing outcome and exposure variables.

Table S4: Power analysis results for the NO_2 exposure: simulated power at specified effect sizes and 1% significance level, given our model and sample size.

	Specified Risk Ratio														
	1.01	1.02	1.03	1.04	1.05	1.06	1.07	1.08	1.09	1.10	1.11	1.12	1.13	1.14	1.15
Diabetes	0.017	0.019	0.035	0.045	0.056	0.070	0.095	0.126	0.177	0.211	0.274	0.336	0.416	0.494	0.585
Heart disease	0.013	0.025	0.037	0.075	0.143	0.216	0.314	0.431	0.542	0.667	0.774	0.879	0.923	0.954	0.986
Hypertension	0.020	0.074	0.208	0.408	0.599	0.828	0.931	0.993	0.999	1.000	1.000	1.000	1.000	1.000	1.000
Stroke	0.012	0.009	0.010	0.012	0.021	0.027	0.042	0.056	0.057	0.074	0.088	0.111	0.130	0.162	0.182
Asthma	0.021	0.030	0.056	0.110	0.193	0.289	0.393	0.548	0.674	0.783	0.867	0.922	0.964	0.986	0.993
COPD	0.010	0.014	0.020	0.036	0.057	0.098	0.137	0.189	0.252	0.342	0.422	0.518	0.604	0.685	0.771

Table S5: Power analysis results for proximity to major and minor roads: simulated power at specified effect sizes and 1% significance level, given our model and sample size.

	Specified Risk Ratio									
	0.99	0.98	0.97	0.96	0.95	0.94	0.93	0.92	0.91	0.90
Diabetes										
Distance to major road	0.035	0.102	0.265	0.457	0.652	0.803	0.900	0.951	0.985	1.000
Distance to minor road	0.043	0.154	0.316	0.534	0.745	0.881	0.972	0.996	1.000	1.000
Heart disease										
Distance to major road	0.028	0.106	0.251	0.436	0.624	0.780	0.892	-	-	-
Distance to minor road	0.048	0.186	0.444	0.723	0.903	0.980	0.996	1.000	1.000	1.000
Hypertension										
Distance to major road	0.146	0.629	0.956	0.998	1.000	1.000	1.000	1.000	1.000	1.000
Distance to minor road	0.142	0.698	0.963	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Stroke										
Distance to major road	0.026	0.036	0.056	0.084	0.126	0.183	0.238	0.294	0.359	0.414
Distance to minor road	0.024	0.046	0.076	0.132	0.198	0.279	0.367	0.458	0.559	0.633
Asthma										
Distance to major road	0.046	0.261	0.611	0.873	0.978	0.999	1.000	1.000	1.000	1.000
Distance to minor road	0.040	0.183	0.495	0.805	0.949	0.995	1.000	1.000	1.000	1.000
COPD										
Distance to major road	0.033	0.142	0.351	0.605	0.802	0.930	0.983	0.994	1.000	1.000
Distance to minor road	0.033	0.125	0.344	0.603	0.823	0.935	0.988	0.997	1.000	1.000

Figure S1: Residential mobility threshold sensitivity analysis for diabetes, asthma, and COPD. Forest plots show estimated effects with 99% confidence intervals and sample size at each threshold.

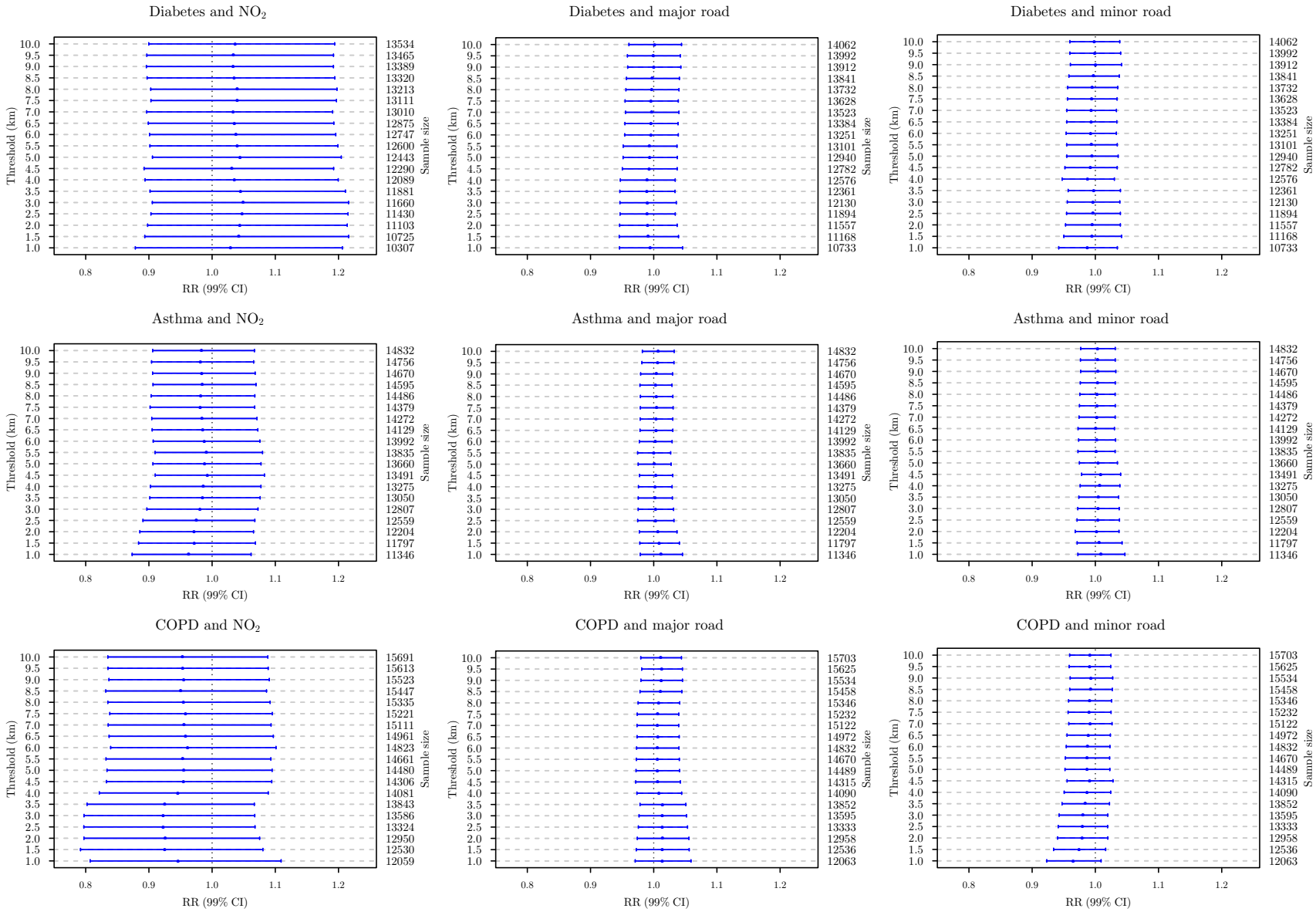

Figure S2: Residential mobility threshold sensitivity analysis for heart disease, hypertension, and stroke. Forest plots show estimated effects with 99% confidence intervals and sample size at each threshold.