PEER REVIEW HISTORY

BMJ Open publishes all reviews undertaken for accepted manuscripts. Reviewers are asked to complete a checklist review form (**http://bmjopen.bmj.com/site/about/resources/checklist.pdf**) and are provided with free text boxes to elaborate on their assessment. These free text comments are reproduced below.

ARTICLE DETAILS

TITLE (PROVISIONAL)	Long-term ambient air pollution exposure and self-reported morbidity in the Australian Longitudinal Study on Women's Health: a cross-sectional study
AUTHORS	Lazarevic, Nina; Dobson, Annette; Barnett, Adrian; Knibbs, Luke

VERSION 1 - REVIEW

REVIEWER	Barbara HOffmann IUF Leibniz Institute for Environmental Medicine and Medical Faculty, University of Düsseldorf
REVIEW RETURNED	29-Jun-2015

GENERAL COMMENTS	In this manuscript, Lazarevic and colleagues report the results of a cross-sectional analysis of the Australian Longitudinal Study on Women´s Health. They analyse the association of long-term exposure to NO2 and traffic at the residential address with various self-reported health outcomes and symptoms. Inspite of a relatively large sample size of approximately 12.000 and 35.000 participants, respectively, they do not find any associations in the adjusted analysis. The research question is not novel, however it is the first report from a largeAustralian cohort and therefore warrants attention. This report also encompasses several interesting aspects of air pollution research. For example, the cohort consists of randomly chosen women from a seemingly nationwide obligatory data base. Many cohorts used in air pollution epidemiology are convenience samples or special population subgroups, limiting their representativeness for the general population. Unfortunately, the authors do not indicate the response rate to be able to judge the representativeness for Australian women. Another nice aspect of this study is the intended overrepresentation of rural populations – most air pollution studies focus on urban populations – making this specifically interesting. I also like the idea of including a "negative control" health outcome, even though it could be an object of debate whether iron deficiency is actually unrelated to chronic air pollution. In this context, I would have favored a "positive control", meaning that it would be nice to demonstrate a relationship between some external factor (possibly related to air pollution) and the health outcomes under investigation. Still another nice aspect of this work is the care that has been taken to exclude misclassification of exposure through relocations in the prevalence analysis. Limiting the study population to those participants who did not relocate during the 3 years prior to the prevalence assessment is a straight forward way to investigate potential misclassification.

	The strenghts of this work are balanced by several important limitations or shortcomings, which decrease the interpretive power of this study and do not make this a convincing null effect paper yet. Major uncertainties result from the self-report of health outcome and covariate data. No validation results are presented, showing the precision of outcome (or covariate) assessment. This would however be very important to convince the reader that the null effect is not due to outcome misclassification. It is also unclear how participants with missings or “do not know” were treated in the assessment phase (was there an option in the questionnaire?) and in the analysis. Missingness could potentially induce a systematic bias because it could be related to exposure (i.e. via education and socio-economic status). If I understood the analysis correctly, this was a full-case analysis, meaning that all observations with at least one missing information were excluded from the analysis. This seems to be the case for more than half of the participants, since the sample size in the health outcome analysis drops to less than 10,000. Have you thought about using other methods to treat missingness, such as imputations, separate outcome categories, etc. With small associations as usually expected in air pollution epidemiology, even residual confounding coming from crude covariate adjustment (age in 3 categories only, BMI from self-report, lack of smoking dose, etc.) might be sufficient to overshadow actually true associations, if they are related to the exposure. A more detailed inclusion of covariates and checking whether the covariates are associated with the outcomes in the expected direction (biological plausibility) would be helpful. Surprisingly, the reported symptoms also show no association with exposure, however the questions might have been too unspecific. Do you have any questions available that compare to standard questionnaires like the ISAAC or ECRHS questionnaires? Another limitation is the restriction of exposure to NO2 and traffic exposure. At the levels encountered in this study, NO2 might not be an appropriate indicator, however, one wonders about the low exposure levels because urban participants where included in the study as well, weren´t they? I could not find a thorough exposure description and discussion on that point, including a description of the study area and its major sources of air pollution, the spatial resolution of estimates, predictors of the LUR model, and a thorough description of (time-varying) population exposures. Other exposures might be more important such as fine particulate matter and soot. It seems like these exposures were not available, which should be mentioned. Also, the analysis of traffic proximity is a bit unclear and potentially does not identify the highly exposed as opposed to participants with a medium or lower traffic exposure. What was the traffic density on the major roads in your study? One further limitation of this study is that it does not make optimal use of the available data. This is a longitudinal study, but the analysis is cross-sectional – even though it is named a relative risk in the table, which is quite confusing. Why not do an analysis of incidence in the first place? This would be much stronger and some of the limitations (i.e. unclear direction of cause and effect) could be remedied from the start. It is also not clear why symptoms are treated in a repeated measures analysis. If symptoms are

	considered early signs of disease, than the first mentioning of symptoms should be used as a surrogate for incidence of disease. If symptoms are used as an indicator of acute exacerbations, then a short-term exposure estimate would be more useful. Some more technical aspects are the inconsistencies of numbers. In the abstract and the methods, 26,991 participants are reported, yet the table for symptoms show an N of 35,000. Are these the number of observations, including repeated measures? This should be clarified.

REVIEWER	Yang Gao Assistant Professor, Department of Physical Education, Hong Kong Baptist University, Hong Kong, China
REVIEW RETURNED	02-Jul-2015

GENERAL COMMENTS	Air pollution and health is an old topic in public health but researchers never lose their passions on revealing their relationship and protecting people from air pollution as no one can live without air. Recently, compelling evidence has shown adverse effect of air pollution even at relatively low level. It is particulate matter (PM, especially PM2.5), not gaseous pollutants, which is thought to be more noxious and play an important role in the adverse effect on health. While nitrogen dioxide (NO2), a gaseous pollutant, is considered a marker for the air pollution mixture of combustion-related pollutants, in particular traffic exhaust or indoor combustion sources. Though acute effect of short-term exposure and chronic effect of long-term exposure on mortality have been well documented, there are still knowledge gaps in the relationship between long-term exposure and morbidity. This manuscript aimed to examine the relationship between long-term exposure to ambient air pollution (indicated by NO2) and several self-reported diseases and symptoms of interest in Australian women, who lived in a relatively clean region, which addresses an important research question. Overall, it is well organised and written well in English. I would suggest you consider it for publication if the authors could address my following comments: 1. Sufficiently discuss on limitations of using NO2 as a marker for ambient air pollution, especially for air pollution from stationary sources. 2. How about the relationship between PM (e.g. PM2.5) and the health outcomes in this population? 3. Could it be possible to consider the influence of traffic volume in this study? 4. Lines 42-46, Page 3: add age range to each cohort. 5. Lines 31-35, Page 6: better to present the distributions of annual mean NO2, distance to a major/minor road with diagrams (e.g. box-and-whisker diagram). Better to give the NO2 concentrations in ug/m3. 6. Consider to put Tables S1, S3 and S4 into the main text. 7. Tables S3 & S4: for categorical variables (e.g. cohort), better to present prevalence rate in each category of each variable instead of proportion; for continuous variables (e.g. annual mean temperature), explain what the figures in the brackets present (e.g. 95%CI, or 99%CI?) 8. Integrate Table S5 with Table 1.

VERSION 1 – AUTHOR RESPONSE

Reviewer: Barbara Hoffmann

Comment 1: Unfortunately, the authors do not indicate the response rate to be able to judge the representativeness for Australian women.

Response: We have now included the following sentence:
"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 were 86% for survey 5 and 83% for survey 6; in the older cohort were 77.4% for survey 5 and 70% for survey 6."

We also direct the reader to a recently published ALSWH cohort profile update for further information on the representativeness of the ALSWH data.[1]

Comment 2: I also like the idea of including a "negative control" health outcome, even though it could be an object of debate whether iron deficiency is actually unrelated to chronic air pollution.

Response: We were not able to find any English-language papers linking air pollution with iron deficiency or anaemia after searching Medline, EMBASE, and ISI Web of Science. While we are aware this does not preclude a relationship, we believe this was the most appropriate negative control available in our data.

Comment 3: In this context, I would have favored a "positive control", meaning that it would be nice to demonstrate a relationship between some external factor (possibly related to air pollution) and the health outcomes under investigation.

Response: Testing the sensitivity of a cross-sectional analysis using ALSWH data to detect a known association would be a valuable exercise. The ALSWH survey data have been extensively studied since 1996 with almost 500 published papers to date.[1] However, few of the studies have considered external or environmental factors possibly related to air pollution. Relationships between physical and mental health, adverse climate events (drought, flood, fire) and indicators of environmental degradation (salinity, land surface temperature, and normalised difference vegetation index as a proxy for rainfall) have been considered but no evidence of associations has been found.[2,3] As such, we have not adopted an external positive control in this case.

Many of the covariates included in the models were associated with the outcomes under investigation. For example, Table R1 below shows the estimated effects for age cohort and smoking status in the adjusted disease prevalence and NO2 models.

Table R1: Effect estimates for a selection of covariates from the adjusted disease prevalence and NO2 models. Relative risks with 99% confidence intervals in parentheses.

THIS TABLE WOULD NOT FORMAT CORRECTLY IN THIS HTML BOX. PLEASE SEE RESPONSE TO REVIEWERS ATTACHED TO THIS SUBMISSION AS A 'REVIEW HISTORY' FILE.

Comment 4: Major uncertainties result from the self-report of health outcome and covariate data. No validation results are presented, showing the precision of outcome (or covariate) assessment. This would however be very important to convince the reader that the null effect is not due to outcome misclassification

Response: The validity of self-report data is of paramount importance. The validity of many of the ALSWH variables have been verified elsewhere. For example, in a comparison of ALSWH self-report

data and hospital records in middle- and older-aged women, Navin Cristina et al. found substantial agreement for diabetes and to a reasonable degree for heart disease and stroke, while hypertension was found to be underestimated by hospital records.[4] The validity of self-reported height, weight, and physical activity in middle-aged women have been assessed and confirmed by other studies.[5,6] We have now presented this information in the discussion section of our manuscript:

"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.[61] The validity of self-reported height, weight, and physical activity have also been assessed and confirmed by previous studies.[62,63]"

Comment 5: It is also unclear how participants with missings or "do not know" were treated in the assessment phase (was there an option in the questionnaire?) and in the analysis. Missingness could potentially induce a systematic bias because it could be related to exposure (i.e. via education and socio-economic status). If I understood the analysis correctly, this was a full-case analysis, meaning that all observations with at least one missing information were excluded from the analysis. This seems to be the case for more than half of the participants, since the sample size in the health outcome analysis drops to less than 10,000. Have you thought about using other methods to treat missingness, such as imputations, separate outcome categories, etc.

Response: Our study was a full case analysis, however only 0.6%–18% of participants were excluded due to missing covariate data between crude and adjusted disease and NO2 models, rather than the suggestion of more than half (see sample sizes in Table 2 and number of missing observations by covariate in Supplementary Table S1). The loss of observations due to missing covariate data were small compared with the combined losses due to lack of exposure data (5.5%–8%) and due to restrictions on residential mobility between surveys (20%–25%). However, we conducted sensitivity analyses on our choice of residential mobility threshold which substantially varied the sample size and our conclusions remained unchanged (see Supplementary Figures S1 and S2).

We also assessed the pattern of missingness in the disease data and the missing covariate values were consistent with being missing at random. For the stroke prevalence and NO2 model (the model with the greatest proportion of participants excluded due to missing covariate data, 18%), we imputed missing covariate values using multiple imputation by chained equations in Stata 13.1. The pooled effect estimate across 20 imputations was RR=0.84 (99%CI 0.62-1.13), which suggests no bias due to the missing covariate data in this model (cf. the adjusted effect estimate reported in Table 2, RR=0.83 and 99%CI 0.58-1.19).

Comment 6: With small associations as usually expected in air pollution epidemiology, even residual confounding coming from crude covariate adjustment (age in 3 categories only, BMI from self-report, lack of smoking dose, etc.) might be sufficient to overshadow actually true associations, if they are related to the exposure. A more detailed inclusion of covariates and checking whether the covariates are associated with the outcomes in the expected direction (biological plausibility) would be helpful.

Response: Covariates were included in models with as high resolution as available. Higher resolution variables were available in some ALSWH surveys, however our analysis involved pooling data across surveys and a common (usually coarser) resolution had to be selected. Although space limitations preclude the presentation of all covariate effect estimates, all authors were involved in carefully checking the magnitude, significance, and direction of the estimated effects. Results for a selection of models and covariates are presented in Table R1 above.

BMI and smoking would most likely be confounders via a common cause of education and income, which we have adjusted for. Interestingly in Australian cities there are many wealthy suburbs in the

inner city with high pollution exposures, and studies in other cities have shown that the higher SES areas have the highest exposure.[7] So overall we feel that such confounding would be weak.

Comment 7: Surprisingly, the reported symptoms also show no association with exposure, however the questions might have been too unspecific. Do you have any questions available that compare to standard questionnaires like the ISAAC or ECRHS questionnaires?

Response: The ALSWH surveys cover a wide range of biological, psychological, behavioural, social and lifestyle factors. Consequently the questions on particular outcomes lack the specificity that may be present in surveys focusing on one area of health, such as the ISAAC or ECRHS. The ALSWH was not conceived as an air pollution study, and as a result the questions are not specific to assessing the health effects of air pollution. We acknowledge this limitation of our analysis in the discussion section.

Comment 8: Another limitation is the restriction of exposure to NO2 and traffic exposure. At the levels encountered in this study, NO2 might not be an appropriate indicator, however, one wonders about the low exposure levels because urban participants where included in the study as well, weren´t they? I could not find a thorough exposure description and discussion on that point, including a description of the study area and its major sources of air pollution, the spatial resolution of estimates, predictors of the LUR model, and a thorough description of (time-varying) population exposures. Other exposures might be more important such as fine particulate matter and soot. It seems like these exposures were not available, which should be mentioned. Also, the analysis of traffic proximity is a bit unclear and potentially does not identify the highly exposed as opposed to participants with a medium or lower traffic exposure. What was the traffic density on the major roads in your study?

Response: Both urban and rural participants were included in the study, however rural and remote areas were over-sampled to ensure their adequate inclusion. Figure 1 in the results section shows the distribution of exposure levels in our study.

The study area encompassed the whole of Australia by design (rather than a single city or state). We therefore capitalized on a recently completed national-scale NO2 exposure measurement model.[8] Detailed descriptions of exposure measurement, study area, sources of air pollution, the spatial resolution of predictors in the LUR model, are provided in Knibbs et al.[8], to which we refer the reader. In built-up urban areas the resolution of predictions is up to ~200 m and is approximately 1 to 1.5 km in more remote areas. We did not have comparable national-scale models for other pollutants; this has now been stated in the methods exposure data section of our paper.

We believe NO2 is an appropriately sensitive proxy of combustion-related emissions as it has been found to be highly correlated with other major pollutants near roads,[9] however we acknowledge that the degree of correlation between pollutants depends on them having analogous sources and that it may not hold at all spatial scales.[10] We agree with the reviewer that the health effects of other ambient exposures are also of interest, perhaps even more so for some of the outcomes under investigation, however only ambient NO2 exposure has so far been estimated at a national scale in Australia.

The road proximity variables in our study are based on a qualitative rather than quantitative assessment of traffic density. For example, major roads were considered those with ‘massive’ traffic movement, day and night, while minor roads included local property roads. Quantitative traffic count data are also not available at a national scale in Australia.

Comment 9: One further limitation of this study is that it does not make optimal use of the available data. This is a longitudinal study, but the analysis is cross-sectional – even though it is named a

relative risk in the table, which is quite confusing. Why not do an analysis of incidence in the first place? This would be much stronger and some of the limitations (i.e. unclear direction of cause and effect) could be remedied from the start. It is also not clear why symptoms are treated in a repeated measures analysis. If symptoms are considered early signs of disease, than the first mentioning of symptoms should be used as a surrogate for incidence of disease. If symptoms are used as an indicator of acute exacerbations, then a short-term exposure estimate would be more useful.

Response: We agree with the reviewer that a longitudinal analysis would have been preferable to a cross-sectional analysis. However, exposure data were only available annually for the years between 2006 and 2011, corresponding to a maximum of two surveys per respondent in the ALSWH. Disease questions referred to occurrence during the 3-year period between surveys. To ensure exposure estimates corresponded to this period, only the second of the two surveys was used for disease prevalence. In contrast, symptom questions assessed occurrence in the last 12 months, so both survey responses per respondent were available for analysis, and thus the symptom outcomes were treated in a repeated measures analysis to exploit the available data fully.

For the disease questions, we chose to model relative risks (risk ratios) directly rather than odds ratios to avoid the rare disease assumption, which may not hold for some outcomes under investigation (e.g. hypertension). Our terminology reflects our choice of methodology: for disease prevalence, we estimated risk ratios via population-averaged Poisson regression models with log-link and robust error variance; whereas for symptom prevalence, we estimated odds ratios via multi-level mixed logistic regression.

Comment 10: Some more technical aspects are the inconsistencies of numbers. In the abstract and the methods, 26,991 participants are reported, yet the table for symptoms show an N of 35,000. Are these the number of observations, including repeated measures? This should be clarified.

Response: The reported sample sizes for the symptom data do include repeated observations. This has now been stated in the table footnotes.

Reviewer: Yang Gao

Comment 1: Sufficiently discuss on limitations of using NO2 as a marker for ambient air pollution, especially for air pollution from stationary sources.

Response: Please see the response to Comment 8 of the first reviewer.

Comment 2: How about the relationship between PM (e.g. PM2.5) and the health outcomes in this population?

Response: Please see the response to Comment 8 of the first reviewer.

Comment 3: Could it be possible to consider the influence of traffic volume in this study?

Response: Considering the influence of traffic volume would be a beneficial extension to our analysis, however because we performed a national study and traffic volume data are patchy even in large cities, this approach was not feasible.

Comment 4: Lines 42-46, Page 3: add age range to each cohort.

Response: We have included the age range in each cohort as suggested. The sentence now states:

“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 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 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 at the later survey).”

Comment 5: Lines 31-35, Page 6: better to present the distributions of annual mean NO2, distance to a major/minor road with diagrams (e.g. box-and-whisker diagram). Better to give the NO2 concentrations in ug/m3.

Response: Box plots of annual mean NO2 and distance to roads have now been presented in Figure 1.

For gaseous pollutants conversion from ppb to µg/m3 is dependent on pressure and temperature. To aid in comparison with published studies that were conducted in µg/m3, we suggest a conversion factor of 1.88 based on 25˚C and 1 atmosphere of pressure.

Comment 6: Consider to put Tables S1, S3 and S4 into the main text.

Response: Table S1 has been included in the main text (now named Table 1). Tables S3 and S4 have been retained in the supplement due to journal restrictions on the number of included tables and figures (now named Table S2 and S3, respectively).

Comment 7: Tables S3 & S4: for categorical variables (e.g. cohort), better to present prevalence rate in each category of each variable instead of proportion; for continuous variables (e.g. annual mean temperature), explain what the figures in the brackets present (e.g. 95%CI, or 99%CI?)

Response: For categorical variables, sample prevalence and sample size in each category have now been presented instead of proportions. For continuous variables the figures represent the 5th and 95th percentiles, and this is noted below the tables.

Comment 8: Integrate Table S5 with Table 1.

Response: Table S5 has been integrated with Table 1 (now named Table 2).

REFERENCES USED IN RESPONSE

1 Dobson AJ, Hockey R, Brown WJ, et al. Cohort Profile Update: Australian Longitudinal Study on Women’s Health. Int J Epidemiol 2015;17:1186–94. doi:10.5588/ijtld.12.0959
2 Fearnley EJ, Magalhães RJS, Speldewinde P, et al. Environmental correlates of mental health measures for women in Western Australia. Ecohealth 2014;11:502–11. doi:10.1007/s10393-014-0966-3
3 Powers JR, Loxton D, Baker J, et al. Empirical evidence suggests adverse climate events have not affected Australian women's health and well-being. Aust N Z J Public Health 2012;36:452–7. doi:10.1111/j.1753-6405.2012.00848.x
4 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. Geriatrics & Gerontology International 2015;:n/a–n/a. doi:10.1111/ggi.12442
5 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. doi:10.1111/j.1753-6405.2008.00305.x

6 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. doi:10.1111/j.1753-6405.2010.00618.x
7 Forastiere F, Stafoggia M, Tasco C, et al. 16847936. Am J Ind Med 2007;50:208–16. doi:10.1002/ajim.20368
8 Knibbs LD, Hewson MG, Bechle MJ, et al. A national satellite-based land-use regression model for air pollution exposure assessment in Australia. Environmental Research 2014;135:204–11. doi:10.1016/j.envres.2014.09.011
9 Beckerman, Jerrett, Brook, et al. Correlation of nitrogen dioxide with other traffic pollutants near a major expressway. Atmos Environ 2007;42:16–6. doi:10.1016/j.atmosenv.2007.09.042
10 Miller, Lemke, Xu, et al. Intra-urban correlation and spatial variability of air toxics across an international airshed in Detroit, Michigan (USA) and Windsor, Ontario (Canada). Atmos Environ 2010;44:13–3. doi:10.1016/j.atmosenv.2009.12.030

VERSION 2 – REVIEW

REVIEWER	Barbara HOffmann IUF, Germany
REVIEW RETURNED	20-Aug-2015

GENERAL COMMENTS	Review for R1 „Long-term ambient air pollution exposure and self-reported morbidity in the Australian Longitudinal Study on Women´s Health” This is the first revision of the submitted manuscript and the authors have answered most of my questions sufficiently. A few points should be made more explicit in the text though, since other readers will want to know as well. The authors clarified the drop in sample size. If I have not overlooked it, please add a statement in the text. I am still concerned regarding possible residual confounding given the coarse categories of covariates. A sensitivity analyses in those surveys with a more detailed specification of covariate data should be done to check, whether the inclusion of detailed data makes a difference. A description of exposure was added in the way of a boxplot. Several data points are located at 1 or even fewer meters from high traffic. How is this possible? Please explain the definition of the road proximity measure and check possible implausibilities. Also, the analysis of traffic proximity across the whole span of distances does not really focus on the highly exposed participants (only within the first 100 m), but rather looks at a more general concept of urbanization. To focus more specifically on highly traffic exposed versus low exposure, it would be more useful to truncate the distances at 300 or 400 m (no additional direct traffic emissions from the respective road in question will be measurable beyond that distance). In the rebuttal it is mentioned that nationwide traffic density data is not available. Does that mean that specific areas have this data? Could a sensitivity analysis or a validation be performed to check whether those you are assigned to high traffic proximity actually have high traffic exposure? If not possible, please comment in the discussion on the lack of nationwide traffic density data and the resulting difficulty in identifying the truly highly exposed individuals. Traffic noise was not included. Please comment in the limitations.

REVIEWER	Yang Gao Assistant Professor Department of Physical Education Hong Kong Baptist University Hong Kong SAR, China
REVIEW RETURNED	20-Aug-2015

GENERAL COMMENTS	The reviewer completed the checklist but made no further comments.

VERSION 2 – AUTHOR RESPONSE

We thank Barbara Hoffmann for her careful and insightful review of our manuscript; our responses follow.

Comment 1: A few points should be made more explicit in the text though, since other readers will want to know as well. The authors clarified the drop in sample size. If I have not overlooked it, please add a statement in the text.

Response: The reasons for the drop in sample size are explained in the methods section (sample selection subsection) and detailed in Supplementary Table S1, which shows the derivation of analytical sample sizes and number of missing observations. We have expanded the following sentence and included an additional reference to Table S1:

"We restricted the sample to those respondents with no missing data on the variables used in the analysis, firstly excluding 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 Supplementary Table S1)."

Comment 2: I am still concerned regarding possible residual confounding given the coarse categories of covariates. A sensitivity analyses in those surveys with a more detailed specification of covariate data should be done to check, whether the inclusion of detailed data makes a difference.

Response: We have used high quality individual-level data and have been able to adjust for a wide array of potential confounders, in our opinion to a greater degree than many comparable published air pollution studies. Where more detailed covariate data existed in some surveys, this usually amounted to only one or two additional categories (e.g. smoking, fruit and vegetable consumption). In some instances categories were combined due to low frequencies (e.g. risky and high risk drinkers were combined; separated, widowed and divorced women were combined), whereas for other variables no greater resolution existed than what was used (e.g. physical activity and the self-assessed ability to manage on income available).

We agree that residual confounding is nonetheless a risk, particularly since the ALSWH was not conceived as an air pollution study and as such not all possible confounding variables were collected. However, we do not believe that additional analyses of subgroups would be useful for the purpose suggested. The ALSWH surveys were each performed on a different age cohort. As a result it would not be possible to determine whether differences in effects are due to differing subgroup characteristics, the effects of the more detailed exposure, or whether the differences are indeed due to better confounding control from the use of the (slightly) more detailed covariate data. Differences may also be masked by lower precision due to performing the analysis on a small proportion of the full dataset. We have now briefly acknowledged this issue in the discussion section:

"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."

Comment 3: A description of exposure was added in the way of a boxplot. Several data points are located at 1 or even fewer meters from high traffic. How is this possible? Please explain the definition of the road proximity measure and check possible implausibilities.

Response: Our road proximity variables measure distances to minor and major roads from geocoded residential addresses. Where an address was not matched exactly to the geocoded national address reference file, the next best match was chosen, for example to the neighbouring building or street, or in rare cases to the postcode centroid. There were 310 individuals in the disease models where distances to roads may have appeared unusually small (less than 1 metre) due to geocoding inaccuracy. Geocoding may also not always be precise: an address may be matched to a point on the periphery of a block of land, for example in front of the driveway, rather than to the midpoint of the block. Distances to roads may sometimes appear unusually small due to this imprecision. After excluding respondents with inexact geocoding matches (the 310 cases mentioned above), there were only 13 respondents with distances to roads of less than 1 metre in the disease models. We assessed the sensitivity of our results to the accuracy of geocoding by repeating the analysis on the subset of respondents with exact matches between respondent addresses and the geocoded national address reference file, and our conclusions remained unchanged.

Comment 4: Also, the analysis of traffic proximity across the whole span of distances does not really focus on the highly exposed participants (only within the first 100 m), but rather looks at a more general concept of urbanization. To focus more specifically on highly traffic exposed versus low exposure, it would be more useful to truncate the distances at 300 or 400 m (no additional direct traffic emissions from the respective road in question will be measurable beyond that distance).

Response: We used a continuous measure of distance to roads that ranged from very high exposure (within 50 metres) to very low exposure (more than 5 kilometres). This measure has the greatest variance and hence the greatest statistical power and generalisability. Truncating or categorising a continuous measure almost always reduces statistical power.

While we agree that no direct traffic emissions from the nearest road are likely to be measurable beyond 300 or 400 m, there is considerable overlap in exposure 'zones' from different roads, particularly in urban areas. This is made more complex by the need to take account of wind directions, which exhibit strong seasonal variability in many parts of Australia. For these reasons, we are reluctant to adopt the suggested approach as it might introduce more misclassification than it removes.

Comment 5: In the rebuttal it is mentioned that nationwide traffic density data is not available. Does that mean that specific areas have this data? Could a sensitivity analysis or a validation be performed to check whether those you are assigned to high traffic proximity actually have high traffic exposure? If not possible, please comment in the discussion on the lack of nationwide traffic density data and the resulting difficulty in identifying the truly highly exposed individuals.

Response: Traffic density data for some city areas may be available from their respective traffic authorities, however this would represent only a small fraction of our overall study area. We believe that traffic density data would not be suitable as a validation variable: like proximity to roads, traffic density data is itself an imperfect proxy for the traffic-related component of ambient air pollution. We thus chose to focus our study on exposure estimates from a satellite-based land-use regression model which incorporates data from several traffic-related variables in addition to reflecting air

pollution from other sources. We have acknowledged this issue in the discussion section:

“While our NO2 data have been validated[19], we were not able to assess the validity of our road proximity variables.”

Comment 6: Traffic noise was not included. Please comment in the limitations.

Response: We have now included the following sentence in the discussion section:

“In addition, we were not able to assess the effects of traffic noise and therefore cannot rule out residual confounding by exposure to noise.”