Response to referee #2

This is an introduction paper on the protocol of the AIRLESS study conducted in Beijing. The overall study design is rigorous in terms of the methods presented in exposure and health outcome measurements. The comparison of the health effects of air pollution in hypertensive population between urban and suburban area is innovative and interesting. The scope to explore comprehensive range of exposure and health outcome metrics is the strength of the study. Despite of this, there are several places in the manuscript that need to be improved by better clarifications of the methodology and preliminary results.

Overall: Thank you for taking the time to provide valuable feedback. This paper presented the methodological framework for the collection of detailed medical biomarkers and exposure estimates as part of the AIRLESS project. The comments are extremely relevant as the reviewer has a clear understanding of current gaps on the effects of air pollution on health and the potential of projects like AIRLESS to address such issues.

Regarding the reviewer's comment, we have added more tables and figures into the results and discussion session. The manuscript has also been restructured and certain sections has been revised accordingly. Please find below the point-to-point response.

Comment 1: There are many air pollutants and health outcome measurements presented in this study. For air pollutants, it is possible that many species may come from the same sources and therefore, highly correlated. The association, if identified, may not directly reflect the true toxicity of health effect for a pollutant but an alternation of source-related effect. It is important to think more in the paper about how to make use of the comprehensive exposure data and propose novel method that can leverage the combination of multiple pollutants' effect in the health analysis. For the health outcomes, the multiple biomarkers from the same pathways may generate an issue of multiple comparison (or not). As this is a methodology paper, it would be helpful to include a discussion on this issue.

This comment hits the nail on the head. We understand it's really hard to differentiate the health impacts of species highly correlated in the outdoor environment due to similar sources (i.e. NO_2 and $PM_{2.5}$ both primarily emitted from traffic). The application of personal monitor and the developed time-activity-location model would be helpful to separate the effect of the key pollutant.

We have re-written parts throughout the manuscript to stress this point. We have modified the background section to stress the four wider research gaps this project aims to address:

a) To investigate the interactive effects of air pollution and hypertension
b) To establish more reliable links between air pollution and health effects by reducing exposure misclassification.
c) To differentiate source-related health effects of air pollution
d) To investigate the underlying mechanism of air pollution on health

We also revised the manuscript in the discussion section to highlight the advantage of the use of personal exposure to find the difference between ambient personal exposure, and the application of time-activity model to potentially separate the health effect of highly correlated pollutant.

"Firstly, the study deployed a state-of-the-art and validated personal air pollution monitor to improve the personal exposure assessment to multiple pollutants. The high compliance rate of the participants with the study protocol highlighted the feasibility of collecting personal exposure data at high spatiotemporal resolution matched with detailed health assessments. The preliminary results highlight a clear difference between personal and ambient exposure driven by individual activity patterns, meteorological factors and the built environment. In line with previous literature, we show the large biases arising from the use of ambient measurements to represent personal exposure in most epidemiological studies, and the potential of novel sensor technologies to revolutionise future human-based studies.

Secondly, time-activity-location patterns of individuals are important determinants of personal exposure but due to the relative difficulty of collecting such information, they have rarely been taken into account by air pollution epidemiology. For the relatively sedentary participants of this panel study, the home environment was the major contributor to overall exposure, and an important modifier of personal concentrations for all investigated air pollutant species. Exposure differences between the two panels were attributed partly to the variation in domestic energy use. For instance, in winter the urban building stock in China relies on centralised gas heating system, while traditional biomass and coal stoves remain the key emission source for heating and cooking in peri-urban areas. However, the exposure variability between participants was larger than the variability between the two groups, stressing the need to go beyond current methodologies to estimate population exposures."

Regarding the examination of health outcomes, we reckon that an increasing number of biomarkers in a study will increase the difficulties in explaining the biological mechanisms, as some of the biomarkers may share similar pathways or be regulated in a more complicated biological network (eg. Cytokines). The fast development of omic-related analysis, which could generate thousands of biomarkers, will be helpful but meanwhile add more challenge in understanding the biological mechanism. We have considered this issue for the analysis of multiple biomarkers, and revised the strategy for analysis accordingly in 2.9 statistical analysis, as followed:

"To examine the effect of air pollutant on multiple biomarkers (e.g. metabolome and transcriptome), false discovery rate (FDR) adjusted p-value will be applied to detect the statistical significance. Pathway enrichment analyses based on the changes in multiple biomarkers will be used to investigate potential mechanisms."

Comment 2: regarding to low-cost sensor technologies, "sensor technology is complex and requires careful calibration both internally within device and externally across the devices and with other standard instruments under various environmental circumstances. The current study only reported the specifications and performances of the PAM monitor, but did not include detailed descriptions on how to ensure the accuracy of the monitors in the real world measurements."

Thanks for the comment from the reviewer. We understand the importance of the validation of the personal monitor (PAM) we used in this study. The performance of the sensors integrated in the PAM has been characterised extensively in a previous publication (Chatzidiakou et al., 2019). We added a paragraph in the subsection of method (2.5 Personal exposure in revised manuscript) to summarize the key results from that paper, detailed as below:

"The characterisation of the performance of the air quality sensors integrated in the PAM is presented in a previous publication (Chatzidiakou et al., 2019). Briefly, all PAMs were calibrated in two outdoor co-location deployments at the urban PKU site next to reference instrumentation for one month after the winter and summer deployments to participants. The performance of the NO_2 and $PM_{2.5}$ sensors was additionally characterised in an indoor microenvironment next to commercial instruments. Overall, the air pollution sensors showed high reproducibility (mean R^2= 0.93, min–max: 0.80–1.00) and excellent agreement with standard instrumentation (R^2>0.84 for all sensors in winter, while

R^2>0.71 in summer). Further work (Chatzidiakou et al., 2020) showed that the error of the PAM was negligible compared with the error introduced when deriving exposure metrics from fixed ambient monitoring stations close to the participants' residential addresses. Hence, novel sensing technologies such as the ones used here are suitable for collecting highly resolved personal exposure measurements in large-scale health studies."

Comment 3: The results are relatively simple. At least, the demographics of the study population and the exposure and outcome measurement statistics are needed, so that it is good for readers to understand the overall differences of exposure and outcomes between the two study sites. The results will also help support the proposed hypothesis of the study.

Thanks for the comment. To better characterise the two panels of participants in this study, we have added a table (Table 4 in revised version). Basic health outcomes (such as BMI, WHR, hypertension status) were also included in Table 4. For the other detailed biomarkers, we plan to included in the following papers. Regarding to the exposure in the two sites, we added two figures (Figure 6 and 9 in revised version) to show the ambient and personal exposures during the campaigns in both sites.

We also revised the manuscript regarding to the results of these analysis, which is described in subsection 3.1 "Demographics characteristics of urban and peri-urban participants", 3.2 "Ambient concentration of $PM_{2.5}$ during study periods", and 3.5 "Seasonal and spatial pattern of the difference between personal and ambient exposure".

Comment 4: As mentioned by the authors, one of the major differences in the urban and suburban sites is the contribution of indoor exposure to the personal exposure and the indoor exposure levels are supposed to be higher. Only the outdoor monitoring sites include detailed air pollutant species as compared to the personal exposure. Thus, the importance of the contributions of the species to the personal exposure seems to be attenuated. Will it be possible to use the GPS data to split outdoor and indoor exposure in the health analysis, so that the comparisons are relatively fair?

Thanks for the comment, this is really an important question to understand the health effect of pollutant species. Currently, personal sensors were not applicable to a wide range of pollutant species, especially considering the performance of measurement at a high time-resolution. The common commercial portable monitors used in most of the epidemiological studies are usually targeted on either particles ($PM_{2.5}$, BC, etc), or gaseous pollutants (NO_2, CO etc). The personal monitor we used was developed ourselves which is unique to include both PM and gaseous sensors in one device. This enables the PAM to measure PM in different size fractions, and four species of gaseous pollutants, which will help us to understand the health effect of the key pollutants. We have updated Table 2 (combined the previous Table 2 and 3 together) in the manuscript to describe the physical and chemical parameters of both ambient and personal exposure measurement.

Apart from that, we have also considered the suggestion of the reviewer to split indoor/outdoor exposure to give a more accurate exposure assessment. An automated model was developed to classify time-activity-location patterns based on parameters collected with the PAMs (GPS, background noise, acceleration)[2], which has been described in our newly published paper. The classifications include core location categories*: "home", "work", "other indoor static", "other outdoor static", " travel"*, as well as activities *"cooking", "sleeping"* and modes of transport *("walk", "cycle", "motorbike", "car/bus", "train/tube").*

New subsection (2.6 The time-activity-location model) was added to highlight this methodological element of the project. We also added a figure (Figure 8 in revised manuscript) and a subsection in result (3.4 an illustrative example of exposure misclassification) to show how we apply the time-activity-location to help understanding the potential sources for personal exposure.

References:

1 Chatzidiakou, L., Krause, A., Popoola, O. A., Di Antonio, A., Kellaway, M., Han, Y., ... & Fan, Y. (2019). Characterising low-cost sensors in highly portable platforms to quantify personal exposure in diverse environments. *Atmospheric measurement techniques*, *12*(8), 4643.