Reviewer's report

Title: The effect of exposure to biomass smoke on respiratory symptoms in adult rural and urban Nepalese populations

Version: 2  Date: 7 July 2014

Reviewer: TQ Thach

Reviewer's report:

My specific comments are:

Page 5, line 10: Please be more specific about “…appropriately adjusted…” for confounders. What does "appropriately" mean in this context?

Page 8 in “Statistical analysis” section: What is the difference between using GEE and robust variance estimation on handling correlated data? Any advantage using robust variance estimation vs GEE approach?

Page 8. What is the actual level of PM2.5 corresponding to 2SD of the arithmetic means?

Page 9, line 1: The “potential risk factor” should be “potential confounder”?

Page 9, line 7: The sentence “Those biomass fuel users were more likely to be … lower income” implied some form of causal association. I suggest to rephrase as “The proportion of current smokers … were higher among biomass fuel users”.

Page 10 last line: “These respiratory symptoms … than average exposures.” This sentence was more likely to pick and choose positive results, not all the “Wheeze” and “Breathlessness” followed this observed pattern.

Page 13, line 8: typo on “Nepal8”, “countries.4 20”.

Page 14, Paragraph 2 last line: Please point out the results on “kerosene”. I could not find them anywhere in the text.

Table 3: Since so many outcomes and exposures were tested, Type 1 error should be inflated considerably. Have you thought about correcting Type 1 error? Why or why not?

Table 3: Please specify the change of exposure for odds ratios, such as “Indoor 24-hr mean PM2.5 per 10 #g/m3 increment”.

Table 3: Please provide more detail event count on each outcome and exposure similar to Table 1. We may be able to have a better understanding whether the subjects are balance in different gender and different exposure.

Table 3: Too many outcomes may confuse readers, it is better to focus on several core symptoms such as wheeze, cough and phlegm. On the other hand, better to point out which "sub-outcomes" was used for discussion. For example, which of the three "sub-outcomes" represented the “Wheeze” in the “Discussion” part, “Ever”, “On most days” or “In the last 12 months”?

Table 3: In Column 4 "PM2.5> 2SD of 24-h mean", some of the confidence
intervals for the ORs are wider than those in Column 3 "Indoor 24-h mean PM2.5". The sample sizes in Column 4 should be smaller than that of Column 3. Any justifications for these wider CIs?

Nepal is situated in region where the climate necessitates open windows and fan ventilation throughout most of the year. Consequently the inhalation of high ambient air pollution is inevitable. So how do the authors deal with this external source of air pollution when measuring indoor PM2.5 concentrations?

Are the burning incense and the use of mosquito coil commonly used in Nepal? If so, how these contributions to the indoor air quality may affect your PM2.5 measurement s? Would these affect your interpretation of the findings in anyway?

In general, some of the discussion part should be moved to the Results Section. For example, in Page 13 “Sleeping in a …..urban population”.

**Level of interest:** An article of limited interest

**Quality of written English:** Acceptable

**Statistical review:** No, the manuscript does not need to be seen by a statistician.

**Declaration of competing interests:**

NO