Author's response to reviews

Title: Particulate air pollution and chronic ischemic heart disease in the eastern United States: a county level ecological study using satellite aerosol data

Authors:

    Zhiyong Hu (zhu@uwf.edu)
    K Ranga Rao (rrao@uwf.edu)

Version: 2 Date: 14 May 2009

Author's response to reviews: see over
Reviewer 1

Minor revisions

1. Discussion of the LISA cluster map and physical interpretation of AOD-SMR variation.

Discussion was added. See bottom of Page 13 through Page 14 (Second paragraph in the Result section).

2. Current MODIS aerosol retrieval uses three channels (blue, red and thermal IR).

Ok. I revised the statement. (Line 7, p7).

3. More detail about the temporal averaging of MODIS AOD is necessary.

See Paragraph 2, p7. A two-year average AOD raster data layer (10 km by 10 km grid) was calculated by overlaying the daily AOD layers and using a GIS cell statistics (mean) function. Data from both Terra and Aqua satellites were used. MODIS AOD data are not available every day due to cloud cover. If a pixel does not contain AOD data on a day, it was not included in temporal averaging calculation.

4. Averaging of MODIS AOD to derive county average AOD.

See p10. The 2003-2004 mean AOD raster grid was first resampled so that each 10 km by 10 km grid cell was subdivided into 10 by 10 smaller cells retaining the original AOD values. The purpose of the resampling procedure was to split the 10 km cell on the county boundary into separate parts for neighbouring counties to achieve higher accuracy of county average AOD calculation. The resampled AOD grid was then overlaid with the map of CIHD rates. A GIS zonal statistical function was used to calculate the mean AOD value for each county. The mean AOD value was calculated by averaging AOD values of all cells whose centroids are within the county.

5. About R square (coefficient of determination)

R² values in both the OLS model and the two spatial regression models represent the actual contribution of AOD. The values do not contain the contribution of the spatial lag and error. Although the spatial models contain lag/error terms, they are not traditional multiple regression models. A low R² value does not necessarily mean weak association. We should look at the sign of the coefficient and p value to judge the significance of the association. All the three models have shown significant association b/w AOD and SMR. As mentioned in the background, there are many risk factors of CIHD. We cannot expect a high R² value in a univariate model. “Many researchers often place undue emphasis on obtaining a high value of R². However, R² alone is not always the best measure of the utility of a regression model…… it can be difficult to get even moderately large R²

**Reviewer 2**

**Major points**

1. About mortality rate smoothing to account for the rate instability due to small numbers.

   See p10. The number of counties in the study area is 2,506. 200 counties with unreliable and suppressed disease data were omitted from the analysis. The number of data points (counties) for the statistical modelling is 2,306.

   Yes, we could smooth SMR using empirical Bayesian smoothing or spatial rate smoothing methods (but smoothing SMR might necessitate smoothing of AOD data). However, the disease data set downloaded from the US CDC WONDER online database originally does not contain specific number of deaths for the counties with small number of deaths. These counties are flagged “unreliable” or “suppressed”.

2. Explain random permutation.

   See p11. A random permutation procedure recalculates a statistic many times by reshuffling the data values among the map units to generate a reference distribution. The obtained statistic calculated based on the observed spatial pattern is then compared to this reference distribution and a pseudo significance level is computed.

   This is similar to playing cards. We have 54 cards. We reshuffle the cards 1,000 times in a random way, which generates a statistical distribution of the order of the cards. In spatial analyses, we can also assign data value of a map unit to another. We re-arrange data values in the geographic space many times. Then compare the observed pattern of a phenomenon to the random distribution to see if the observed pattern is random or clustered.

3. Confounding risk factors were not included in the model due to data unavailability.

4. We have examined the scatter plot (SMR-AOD). It shows a linear trend. With many risk factors, we cannot expect high R2. See #5 in responses to Reviewer 1.
Minor points

1. Explanation of the constant rho was added. See p12.

Discretionary revision

Explanation of higher correlation of AOD-PM$_{2.5}$ was given at page 14 and also in the citation [25].

Adjusted made in response to the editor’s request:

All adjustments have been made.