Reviewer’s report

Title: Association between Paraoxonase gene and stroke in the Han Chinese Population

Version: 2 Date: 23 May 2012

Reviewer: Tom Trikalinos

Reviewer’s report:

I read the paper with interest. I have the following comments.

MAJOR COMPULSORY REVISIONS

1. In epidemiological terms, the study is not a cohort. You should call the study a case-control.

2. Was the genotyping 100% successful? This would be quite unusual. What was the number of successful calls per SNP and outcome?

3. The analysis is not clear. Please explain what you did and do not only cite software. More specifically:

A. Why did you adjust for covariates? Due to Mendelian randomization the genetic factors cannot possibly be confounders (please refer to the epidemiological definition of a confounder!). Please drop all adjusted analyses, or if you opt to keep them, provide an epidemiologically convincing explanation for your analytic choices.

B. Is the contrast always referring to the frequency of the minor allele? Please provide a Table with the genotyping results in cases and controls for all outcomes.

C. For haplotype analyses, explain how you reconstructed haplotypes (which method?). In the statistical reconstruction of haplotypes we obtain for each person a vector of probabilities corresponding to each possible haplotype. How did you proceed from this stage onwards? Explain whether you used the probabilities for reconstructed haplotypes in the statistical analysis, or whether you rounded them up in some way. Also, how did you handle unsuccessful genotyping calls?

D. For haplotype analyses, describe how you performed the association analysis, i.e., exactly which statistical analysis the SHEsis (?) software does. Because all haplotypes are (should be!) considered jointly, I was expecting to see an omnibus test for including vs not including the haplotype information in a logistic regression; This could be a joint hypothesis Wald test, or a likelihood ratio test of nested models. At any rate, I was expecting a P-value. The ORs (presumably from coefficients of a logistic regression) refer to an arbitrarily coded reference
category and are not useful unless we see all of them for all analyzed haplotypes, and have their covariance matrix as well. Please amend the description of the methods, and provide complete reporting of the analyses.

E. In Tables 2, 3 & 4, what is the "T-statistic OR"? Is this an unadjusted OR? If yes keep this column and drop the adjusted analyses as per comment 2A.

F. In Tables 2, 3 & 4 list the numerators and denominators for the comparisons in each row. For example in Table 2, additive model you should list the number of minor alleles over the total allele pool (- the unit is the allele -- the denominator should be 2*500=1000 alleles -- or whatever the rate of successful calls was). For the other models the unit is a person, so the denominators should be 500. Do not list the minor allele frequency, it is not useful. People who do meta-analyses would need that information to include your study.

G. Contextualize this study given previous meta-analyses in the field. One I am aware of is PMID: 20856122, and studies 2 of the SNPs you included.

H. Please drop the p-values from Table 1. There is no reason for statistical comparison between the cases and the controls.

I. Please explicitly mention the family of the comparisons that have been Bonferroni adjusted (i.e., adjustments for how many multiple comparisons?). Please explain which P-values have been adjusted (e.g., all, including haplotypes?; only the SNP ones?)

MINOR REVISIONS

J. Change "multivariable test correction" to "multiple comparisons adjustment" in the middle of the paragraph "Single site association" in the Results.

**Level of interest:** An article whose findings are important to those with closely related research interests

**Quality of written English:** Acceptable

**Statistical review:** Yes, and I have assessed the statistics in my report.

**Declaration of competing interests:**

I declare that I have no competing interests.