Reviewer's report

Title: Prediction of lung cancer risk in a Chinese population using a multifactorial genetic model

Version: 2 Date: 6 June 2012

Reviewer: Ruth Pfeiffer

Reviewer's report:


The paper reads well, but there are still several issues that need to be addressed regarding the description of the method. As some aspects of the methods are not clear, it is not possible to assess the results based on them.

Major Compulsory Revisions

Abstract:

The authors should use AUC consistently (not c-statistic, as this term was not introduced)

The authors should give 95% confidence intervals for the UAC value.

The authors should present the UAFC for smoking alone and then for smoking - SNP score in the abstract.

Methods:

page 7: missing genotypes were assigned the value of a heterozygote. Why not use the expected value of the genotypes, i.e. E(X) = 1p?

page 7: There needs to be more detail regarding the bootstrap procedure, was the model refit to the individuals in the bootstrap sample and then tested on the out-of-sample subjects? If the model was not refit in every bootstrap iteration, the estimate of the AUC is overly optimistic. In that case only the results from the second validation should be provided.

Results:

page 8: how many individuals had missing genotype information on each of the 4 NSPs?

Page 9/10: please give the AUC value of the model that only has smoking and the model that has smoking and SNPs score in the test. Please also provide a
p-value for difference of AUC values for the smoking and smoking + SNPs model

Minor Essential Revisions

Please provide allele frequencies for Table 1.

Please better label the results in Table 3: better explain what covariates entered the model for the respective AUC calculation.

Level of interest: An article of importance in its field

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