Reviewer's report

Title: Rheumatic Heart Disease in Uganda: The Association between MHC Class II HLA DRB1 Alleles and Disease: A Case Control Study

Version: 1
Date: 29 January 2014

Reviewer: Armando Pucciarelli

Reviewer's report:

Commentary by Armando Pucciarelli.

This is a very interesting study, however it shows some aspects which need to be better highlighted.

--Abstract, Results. (see also Results in the text of manuscript). "The mean (SD) and median (IQR) age in years for cases and controls were 29.6 (10.2) and 29.8 (18), and 29.8 (9.8) and 31 (20), respectively".

The representation of the data regarding age does not appear to be rational. In fact, in common usage, the data of a given group are represented as mean ± standard deviation if the distribution of data drawn from the study population is normal. Instead, data are usually presented using median plus interquartile range if they are distributed not symmetrically around the mean. For identifying a non-Gaussian distribution of data, the common statistical programs offer the opportunity of exploiting various tests, such as D'Agostino-Pearson method or the Kolmogorov-Smirnov test. Thus, I don't understand why the authors have decided to use both mean and the median for representing the variable "age" in the groups of cases and controls. Please specify the reasons for describing age with mean and median, whose simultaneous use seems redundant to me. Alternatively, I recommend to execute a test to verify the possible normality of data distribution. In case of normal distribution, please report just the mean ± standard deviation; on the contrary, if at least one of the two groups is characterized by asymmetric distribution around the mean, then please adopt the median plus interquartile range to describe the age in each of the two groups (cases and controls).

- Both in the abstract and in the Results section within the manuscript, the representation of the interquartile range is approximate. In fact, when you use the median to describe a continuous variable, the choice of reporting only the value of the interquartile range it is not sufficiently informative, i.e., the simple difference between the value of the 75th and the one of the 25th percentile doesn't suffice, but rather the numerical values of the 25th and 75th percentiles are required. Thus, please eliminate the value of the difference you reported in round brackets, and replace it by writing the median value followed by (IQR = Q1 to Q3).

- Introduction: "Group A streptococcal infection". Please add the term beta-haemolytic, so as to write: "Group A beta-haemolytic streptococcal
infection”. Please always use the adjective “beta-hemolytic” to name this streptococcal bacterial strain throughout the entire text.

- In the Discussion: Rheumatic heart disease could thus depend not only on genetically determined disease susceptibility, but also on the regional and temporal prevalence of a particular streptococcal strain(s). In fact, at least 15 streptococcal strains, each with distinct M proteins, have been associated with outbreaks of rheumatic fever [in this regard see Carlquist JF, et al. (J Am Coll Cardiol. 1995 Aug;26(2):452-7)]. Each different rheumatogenic streptococcal strain might select for a particular DR allele(s) in the pathogenic process, resulting in a high proportion of patients positive for that particular allele during an outbreak. This could contribute to the explanation of differences between studies which would be otherwise difficult to be justified. For example, there is a strong discrepancy between the data exhibited by the present study, which would show some protective role played by HLA-DR1 against rheumatic heart disease (i.e. rheumatic endocarditis in the large majority of cases), and those reported by Carlquist et al. in their meta-analysis (J Am Coll Cardiol. 1995 Aug;26(2):452-7). According to the latter HLA-DR1 is associated with a greater susceptibility to rheumatic endocarditis in some populations, such as black South Africans and Afroamericans, who are phenotypically quite similar to individuals recruited in the present study. Please comment on this point, i.e., the very evident discrepancy between the protective value of HLA-DR1 against rheumatic endocarditis that you have found in your study and the somewhat conflicting value of risk factor found for this parameter (positivity for HLA-DR1) by Carlquist et al. in similar populations consisting of black South Africans and Afroamericans.

In the Discussion,” Identification of individuals with susceptibility gene to RHD could potentially provide an opportunity to screen anyone with streptococcal pharyngitis and place them on primary penicillin prophylaxis. This will make it more cost-effective to provide primary prophylaxis”. It seems to me that the Authors conceive a possible scenario of periodic prophylactical use of benzathine-penicillin or other antibiotic agents to be implemented only for subjects with ascertained genetic predisposition to the rheumatic endocarditis. However, the project should be better clarified. Please provide more information about this project.

- In my opinion, the Authors should better specify the candidate individuals to be proposed for periodical administration of antibiotics active against” Group A beta-hemolytic streptococcus”. In particular the Authors should declare whether this prophylaxis, that is likely to have to be maintained for long period (several years), should be denied to children and adolescents who are negative for HLA DR 11 even though they present with a clinical history of recurrent throat infections by group A beta-hemolytic streptococcus. Please discuss this issue.

**Level of interest:** An article of importance in its field

**Quality of written English:** Acceptable

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

I declare that I have no competing interests