3 - faddoul

Review for the paper "Added Value of Anti-CD74 Autoantibodies in Axial SpondyloArthritis in a Population With Low HLA-B27 Prevalence".

The authors of the manuscript propose using IgG4 anti-CD 74 as a diagnostic tool for Axial SpondyloArthritis in a population with Low HLA-B27 Prevalence. The added value is due to the fact that the prevalence of HLA-B27 in European populations, where it is commonly used as a predictive marker of SpondyloArthritis, is much higher than its prevalence for middle eastern populations.

In what follows are listed some remarks and suggestions related to (i) the scientific content and (ii) the presentation of the paper.

Remarks and suggestions related to the scientific content of the paper

1- The authors propose IgG4 as a diagnostic marker for early SpondyloArthritis. Hence they use Positive Predictive Values (PPV) and Negative Predictive Values (NPV). However, it must be noted that these statistics are not suitable for the kind of study and sampling that the authors describe in the paper. In fact, these statistics **are to be used in cross-sectional sampling**, where a random or consecutive sample of the population is selected irrespective of disease or test status, because PPV and NPV depend on the prevalence of the disease. They **should** not be used in a case-control study (Naeger et al. 2013). In a case-control study, sensitivity and specificity are more suitable, however they do not convey exactly the predictive power of the test.

2- The authors use the Area Under the Cure (AUC) to assess the quality of IgG4 as a discriminating marker for SpondyloArthritis. Indeed the mathematical interpretation of AUC is that it is equal to the probability that the test result measured in a randomly selected D^+ person is higher than that measured in a D^- person (Altman and Bland 1994). While, such interpretation is valid for unmatched case-control studies, **it does not hold for <u>matched</u> case-control studies** (Brentnall et al. 2015).

3- The authors state that the optimal cut-off of the ROC curve was 0.5148. However, they do not mention **which criterion they chose to determine such optimality**! There is no single optimum fits all criterion. In fact the optimality criteria depends on the problem at hand were a specific compromise is usually done between sensibility and specificity. Some commonly used criteria can be found in (Metz et al. 1978; Sox et al. 2013).\

4- The main rational for conducting this research is that HLA-B27 is used as a predictor for SpondyloArthritis in European populations where its prevalence is higher than in Middle Eastern populations. I **assume** that **although HLA-B27 prevalence is lower in Middle Eastern populations the prevalence of SpondyloArthritis is not lower in the same ratio than that of European populations.** As such the predictive powers of HLA-B27 is insufficient in Middle

Eastern populations. Indeed, if the prevalence rate ratio of SpondyloArthritis is equal to that of HLA-B27 than this study would be pointless. The authors do not explicitly state and analyze this necessary premise of their work. They do not even **state explicitly** the prevalence of SpondyloArthritis in Middle Eastern population and in European populations.

5- The use of statistical inference assume rigorous probabilistic sampling. I acknowledge that such endeavor is more easily said than done, however, even in cases where such sampling was not done, it must be clearly stated in the limitations of the study.

6- Statistical significance (p-values, etc.) and effect quantifications (AUC, sensitivity, specificity, etc.) are used arbitrarily in the paper.
A p-value represent the probability that the observations are due to chance given a null hypothesis. For example, $p = 0.008$ was calculated for the fact that AxSpA patients were slightly older than BD (34 and 30 years). What is the null value of this p-value? In fact such a p-value might, for example, indicate the propensity of young people to participate in scientific studies. Such conclusion does not fit into the stated objectives of the study.

7- Last but not least, it appears from the attached documents that data manipulation was done extensively to palliate for several failings in the study process:

   a. Confusion of the sera during transportation to Germany;
   b. Inexplicable abnormal differences between mean measurements of the first batch and the second one;
   c. Change of the testing technique;
   d. Mixing of the results of the two batches without any scientific justification.

I believe that such deficiencies in the study process raise very serious doubts about the validity of the results. In any case, even if the authors chose to go ahead and publish their results, at the very least they must have described these limitations and shortcomings in the paper. Failing to do so might expose this publication and related papers to suspicion.

Remarks and suggestions related to the presentation and format of the paper

The paper is concise well presented and easy to follow.

Rafic FADDOUL, Ing. Ph.D.
Faculté d'ingénierie – Ecole Supérieure d'Ingénieurs de Beyrouth: ESIB
Université Saint-Joseph de Beyrouth USJ – LIBAN
Campus des Sciences et Technologies CST, Mar Roukos, Dekwaneh
B.P. 11-514, Riad El Solh, Beyrouth, 11 07 2050 LIBAN
Tél : +961 3 847093
Email: rafic.faddoul@usj.edu.lb

Rôle = Dr. de statistiques aussi.

References:

Naeger, D. M., Kohi, M. P., Webb, E. M., Phelps, A., Ordovas, K. G., & Newman, T. B. (2013). Correctly using sensitivity, specificity, and predictive values in clinical practice: how to avoid three common pitfalls. *AJR. American journal of roentgenology*, *200*(6), W566-W570.

Altman, D. G., & Bland, J. M. (1994). Diagnostic tests 3: receiver operating characteristic plots. *BMJ: British Medical Journal*, *309*(6948), 188.

Brentnall, A. R., Cuzick, J., Field, J., & Duffy, S. W. (2015). A concordance index for matched case–control studies with applications in cancer risk. *Statistics in medicine*, *34*(3), 396-405.

Sox HC, Higgins MC, Owens DK, editors. Medical Decision Making. 2nd ed. Oxford, UK: Wiley-Blackwell, 2013. http://dx.doi.org/10.1002/9781118341544

Metz CE. Basic principles of ROC analysis. Semin Nucl Med. 1978;8:283–98. 10.1016/S0001-2998(78)80014-2