BAISHIDENG PUBLISHING GROUP INC
8226 Regency Drive, Pleasanton, CA 94588, USA
Telephone: +1-925-223-8242 Fax: +1-925-223-8243
E-mail: bpgoffice@wjgnet.com http://www.wjgnet.com

Reviewer's code: 03261349

The aim of the paper by Smarrazzo et al was to evaluate the variability of different ELISA kits for antitransglutaminase assay in different European countries. The main drawback is that a gastroenterology audience may not be attracted by this specific research. Indeed, Authors have planned this study in a clinical pathology/laboratory medicine perspective, albeit the results may be very important for clinicians. Further concerns are:

1) A linguistic revision is needed. The final paper was revised by all the authors (among which many have an international experience on this subject) and, finally, by an external service (American Journal Experts), as could be seen by the certificate attached.

2) The results section is a bare comment of figures and tables. It should be more descriptive. The main conclusions obtained by the results are resumed at the end of the section.

3) Bibliography is too scarce and inadequate to support the discussion. A systematic review of the literature did not evidenced any significant research other than the studies already reported in the reference section (key words used: Transglutaminase, quality, control); the critical point of this subject is the absence nowadays of similar studies about the quality control of transglutaminase testing; most of the literature is pointed on the research of new epitopes of transglutaminase or different antigens for the diagnosis of celiac disease like biomarkers of intestinal damages instead of investigating the commonly used diagnostic tests

4) Authors did not evaluate the distance between the central laboratory and the other centers, nor the time elapsed for the delivery of vials. Such factors could interfere with the final results (e. g. due to antigen degradation).

All the vials were delivered at the same time to all participants at the end of our last meeting in Palermo. The vials were stored in small polystyrene-boxes and refrigerated at 4°C. All the participants returned to their laboratory in 12-24 hrs ensuring the stability of the samples

Reviewer's code: 03309172

The authors have evaluated the accuracy of various commercially available TGA kits across different mediterranean countries. Knowing that TGA values can be determined accurately is important for the use of TGA testing in CD diagnosis, and the study therefore provides useful information for the evaluation of clinical guidelines. The possibility of basing diagnosis of CD on TGA testing without the need for endoscopy is intriguing. However, the study does not say anything about how accurately TGA values reflect clinical disease, only how accurately TGA values can be determined in different settings. It is therefore slightly misleading when the conclusion states: "This study demonstrates that TGA titres more than 10 times higher than the cut-off value are reliable, so clinicians can be confident in establishing a diagnosis on the basis of the TGA assay without using invasive techniques." Also in the discussion: "Thus, TGA could be considered a reliable basis for the diagnosis of CD, in light of its robust estimation across Mediterranean countries." I suggest that the authors revise the text and make it clear that the predictive value of TGA testing for celiac disease is not investigated. It should also be mentioned in the text that the official guidelines so far only

recommend serology-based diagnosis of CD in children. As many cases of CD are discovered in adulthood, the possibility of basing CD diagnosis on TGA testing also in adults would broaden the implications of the study. The aim of our study is to reinforce the confidence of clinicians about TGA assay and secondarily to reinforce the role of transglutaminase in the diagnosis of celiac disease. Other studies had already evaluated the correlation between the presence of high titers of transglutaminase and villous atrophy, suggesting the diagnosis of CD omitting duodenal biopsy (ESPGHAN criteria 2012). Some studies, like that of Beltran L. et al, reported a direct correlation between TGA levels and Marsh degree, even if this correlation is not yet received in the ESPGHAN guidelines.

Minor comments: In some of the graphs, there are circles or other symbols accompanied by a number. What do these numbers reflect? The symbols present in some graphs represents the outlines data (values interpreted as abnormal by SPSS, the program used for the statistical analyses); the near number are the ID number of the SPSS table.

In the discussion it says: "a minute variability in the starting sample, before dilution (for example, estimated value 120 versus true value 115) would have produced a progressive increasing error of estimates across dilutions, but again, this is not relevant to our study." I do not understand this argumentation. The deviation should remain the same when you dilute the sample. It was just a simple consideration: the general underestimation observed in the study could be a "false underestimation", perhaps due to an error of dilution during the preparation of the vials by the manufacturer; this error would increase in the subsequent dilutions, with the result of the false underestimation. In few words, maybe 8 different laboratories tells the truth against just one laboratory (the manufacturer). But, as said, this is not relevant to our study, because the specific results do not interfere with the overall accuracy and precision.

Reviewer's code: 03260134

This is one of few study that aims to estimate the precision and the accuracy of the TGA assays used in 7 Mediterranean countries. The results of the study are supportive of the new ESPGHAN guidelines concerning CD without a biopsy in some cases. The aim of the study is clearly fulfilled. Could the authors please clarify the following issues:

1) Page 5, paragraph 2: what were the concentrations of vials S1-S4? Could the authors give the exact numbers as they did with S5 vial? We would precise that the concentrations of each vials were known only by the manufacturer, and communicated only after the conclusion of the study. By the way, these are the values: S1=0 S2=10, S3=20, S4=50, S5=100, S6=<5 (negative control), S7=40-80 (positive control).

2) Table 1: What is the difference between the two kits used in Greece and Tunisia? They seem to have the same origin (Inova Diagnostics) but different cut-off. Could the authors explain this fact? these kits were produced by the same manufacturer, but they are different products (Tunisia: QUANTA Lite® R h-tTG IgA ELISA; Greece: QUANTA Lite® h-tTG IgA ELISA). On the website of the Inova Diagnostics could be found the technical schedule of both kits, in which it can be seen also

the different cut-off.

3) Page 6, paragraph 3: at what timepoint was We do not understand your question.

4) Table 2: Could the authors note in table 2 which of the expected values is what? S1-S5? What about the Negative and positive controls? We ordered the samples in a crescent degree of concentration, not according the number of the vials. By the way, these are the values: S1=0 S2=10, S3=20, S4=50, S5=100, S6=<5 (negative control), S7=40-80 (positive control).

5) Page 10, last paragraph: do the authors have any possible explanation about the greek results? Why Greek centre constantly shows values above the expected? As said in the "Methods" section, all the analyses were manually made, and all participants were invited to run the samples directly without any dilution. This process was the preferred way to avoid variability due to automatic dilution using mechanic dosage of anti-transglutaminase; on the other side, this could expose to human error during the analyses, and it must be take into account in the interpretation of the results.

6) Do the authors have any suggestions about how we could increase the reliability of TGA assays especially at low TGA level? At this moment, the best way to have an higher reliability of TGA assays is to repeat the test; a second test that confirms our previous result will reinforce our suspicion. As discussed also in the final section of the study, the general underestimation observed could suggest that a borderline result could be, instead, positive, so this is another reason for a retesting. Evermore, there are some research group actually working on the so-called "tTg-neo", that seems to have high sensitivity and specificity, but their use in clinical context is still poor.