May 15, 2013

Editors,
BMC Medical Research Methodology

Dear Editors,

We are herewith enclosing our revised manuscript along with point by point responses to the reviewers’ comments. As per the instructions, we have added competing interests and authors’ contribution sections in respective slots. The references styles are now formatted according to the journal style. Please let us know if you have any further questions.

We thank you for your consideration of our manuscript.

Sincerely,

Shrikant I. Bangdiwala
Responses to reviewers

**Reviewer: Guanmin Chen**

*General comments:*
The author reported a simple method to visualize the agreement between two raters using the agreement chart. General, this is a nice short report to illustrate useful tool for the agreement between rates. The agreement chart is a useful complement to present the agreement.

Thank you.

*Major Compulsory Revisions:*
No.

*Minor Essential Revisions:*
1. The title is too short and simple

We purposefully kept it short and simple for readers to remember the title easily.

2. Please input more details for constructing agreement chart.

Section 2 of the manuscript provides the details, plus a software reference is also provided. A reference is given [Muñoz and Bangdiwala (1997) *J Applied Statistics*) that also provides the details.

3. Please discuss the advantage and limitations for agreement chart.

A new paragraph has been included in the Discussion (Page 8, line 16) addressing advantages and disadvantages.

4. Please add more on how to use the agreement chart to evaluate the agreement.

A new paragraph has been included in the Discussion (Page 8, line 22) addressing usefulness in evaluating agreement and disagreement.
Reviewer: Miao-Yu Y Tsai

General concerns

1. I agree that graphs can provide more visual impression than that of summary statistics. Besides, the agreement chart is used to be a complement to agreement measurements for ordinal data rather than an alternative. However, the visual representation scheme for comparison of concordance is subjective and arbitrary as the authors said. Furthermore, not only the order of categories has much influence on the agreement chart but also the sample size and the number of categories. Therefore, the agreement chart can’t offer a reliable technique and may lead to obtain inconsistent and unstable interpretation. This appears to be an important drawback that is not an inherent problem in some existing approaches.

The objective of the agreement chart is as the reviewer mentions - to provide a complementary visualization to agreement measures. The subjective nature of visual assessments is carefully explained by the provision of guidelines for interpretation of various charts. Note that the chart is NOT designed for nominal categorical variables but, as originally mentioned in the manuscript, for ordinal level categorical variables so that the order of categories is fixed – we agree with the reviewer, and that was already stated in the manuscript in its original submitted version. The articles by Muñoz & Bangdiwala (1997) *J Applied Statistics* and Shankar & Bangdiwala (2008) *J Applied Statistics* shows the stability of the B-statistic derived from the agreement chart with respect to different numbers of categories. We thus disagree with the reviewer’s statement that the agreement chart is not reliable. There also are no graphical ‘existing approaches’ for visualizing agreement.

2. The article only provides the construction of the agreement chart and three examples for illustration. The contribution of the article is less. More research results such as development of agreement methodology and comparisons between different methods would have helped.

The agreement methodology and comparisons of the B-statistic with the well-known Cohen’s *kappa* measure of agreement have already been presented in the already cited Muñoz & Bangdiwala (1997) and Shankar & Bangdiwala (2008) *J Applied Statistics* articles. The objective of this manuscript is to provide examples of the utility of the agreement chart for non-statisticians. We have added a reference to the previous articles for those readers interested in the methodology.

Specific comments

1. Abstract is missing.

The abstract was present but was called Summary. We have changed the label to Abstract.

2. On page 5. The value of weighted *kappa* for Tabár scale is lower than that of unweighted *kappa*. Is it a typo?

It is not a typo. We checked the original manuscript – Table 2 of Garrido-Estepa et al (2010) *BMC Cancer*, and the numbers are correct. It can happen depending on the chosen weights for the weighted *kappa* and in this example both absolute and quadratic weighted *kappa*’s produce smaller estimates than the unweighted *kappa*. 
3. **On page 6, -line 5. Who is Rater B?**

We have removed the label ‘Rater B’ and clarified who the person is as ‘the Winnipeg neurologist’.

4. **On page 6, -line 5 to -line 1.** The two sentences “In Figure 1 we first note that Rater B has a bias towards ‘certain’ and ‘probable’ since the ‘path of rectangles’ is above the diagonal line of no bias in both charts. Furthermore, Rater B tends to be even more biased towards the first two categories for his/her patients (Figure 1A), leading one to guess that the masked evaluation of patients was not truly masked.” would lead to a misunderstanding. The departure of the path of rectangles from the 45° diagonal line implies that the two observers tend to rate different categories for patients. Thus, using “observer bias” to describe the phenomenon is not appropriate.

The correct use of ‘observer bias’ was used – ‘bias’ in studies of concordance/agreement refers to the situation when the marginal distributions of classifications of the observers are different. This is reflected in the agreement chart by the ‘path of rectangles’ – obtained by joining the vertices of the rectangles in the chart – deviated from the 45° diagonal line.

5. **Discussion (Section 4) contributes less and should be rewritten.**

The Discussion has been extensively revised following the suggestions of the reviewer.

6. **Check the style of all references.**

References are formatted to *BMC Medical Research Methodology* style. The problematic style is for reference #9, the `vcd` package in the software R, but it has been fixed.