Response to Reviewers
Manuscript # NRES-D-08-00245

Title: Use of the Concordance Correlation Coefficient when Examining Agreement in Dyadic Research

Please extend the authors' appreciation to the editor and two reviewers for their thoughtful responses to this manuscript.
Revising the manuscript according to their comments has added considerable strength.

Response to Editor

Page #	Para #	Line #	Table #	Comment
3-4	2	14-18	Table 3	Editor: please give special attention to showing results for the three statistical tests that you discuss and explain why the kappa statistic was not selected for use. Response: A revised Table 3 now contains results from all three statistical tests. Table 4 was previously Table 3. The kappa statistics for ordinal data depends on the choice of weights. A new paragraph has been added along with references describing why Kappa was not selected for use in this analysis.
				Editor: TEXT -- Please use Times New Roman font with 12 point pitch. FIGURE 1 -- Use x & y axes only (remove top and right side lines) Response: The font has been corrected throughout paper and in the Tables. The Figure has been corrected as requested

Response to Reviewer #1

Page #	Para #	Line #	Table #	Comment
3 4	2 1	14-18 17		Reviewer: why do the researchers want to evaluate congruence of symptoms between patients and family? Are they interested in inter-rater reliability? Do they want to know if there are statistical differences in ratings on which to make clinical decisions? Curiously the kappa statistic (Landis & Koch, 1977) was not used. If the goal is to evaluate inter-rater reliability-this more commonly used statistic would control for expected agreement between symptoms that are frequently reported such as fatigue and dyspnea. The Phi statistic has also been used to evaluate congruence between raters. **Specific recommendation*: P6, L9-10-Agreement between 2 observers is usually evaluated via the PCC or ICC- as above (did you consider using the Kappa or Phi statistic?). Response: A paragraph has been added describing why Kappa or Phi was not used in this analysis. References have also been added.
				Reviewer: The authors state that "the ICC and PCC are used to

4	1	7		examine if measures vary in the same direction." P4, L2-3. However, the PCC measures the strength of an association, whether it is in the same direction or not (i.e. positive or negative). So the case for using PCC or CCC remains unclear. Response: In this study we are looking at cases where we expect both scores to vary in the same direction. This has been clarified in the introduction.
6	2	13	Figure 1	Reviewer: The authors go on to state that "the CCC was developed and tested in the biopharmaceutical industry to assess agreement between methods of measurement" (P4, L8-9). This implies that it is used in the case of multiple measures. For example, there are various methods to test whether a drug is "effective" including physiologic measures, serum drug levels, or biomarkers. Correlations would yield some information about the relationships of the measures but that is not the same as examining congruence between two raters administering the same measure. One would hope to achieve perfect inter-rater reliability but would not expect perfect correlations between different measures. Response: In this paper, we avoid the concept of 'reliability' because it sometimes means agreement and sometimes it means correlation. This issue is addressed in the first example.
				Reviewer: Results- No symptoms had good agreement between patient and family. This seems like a potentially serious problem

9	1	3-10		particularly if the family member is serving as a proxy for the patient. Please address this. *Specific recommendation:* Please state why the examination of congruence in symptoms between patient and family is important. (P9, L14-16 describes what you are interested in-absolute agreement-but not why). Response: The concept of congruence is important within families as the family member is usually involved in the assessment of their partner. It is important to learn which symptoms are more difficult for family members to recognize in order to begin to educate them on more subtle signs of heart failure. This is addressed at the end of the methodology section. Complete details about this study will soon be published in the databased article currently undergoing its second review. This reference has been added.
11	1	3-13		Reviewer: The discussion of the coefficients is thorough. When looking at the findings in Table 3. It is apparent that the symptoms with the least agreement are probably the symptoms that are most frequently reported hence (as with the Kappa statistic) the findings are not unexpected. *Specific recommendation:Could you address this? Response: Many of the symptoms that were in least agreement are more subtle and although may be reported by the patient, the family member may not notice unless they learn the cues to look for or validate with the patient. This statement has been added to the discussion section.
				Reviewer: *Specific recommendation*: P3, L 20- Recommend Pearson's r coefficient (lower case) or simply Pearson's correlation

4		1		coefficient. Response: This has been corrected on page 4 and throughout paper.
8	1	9		Reviewer: *Specific recommendation: P8, L16-17-If the CCC requires that score differences between observers be small, won't an 11-point scale result in potentially large differences? Response: Instead of saying that it (the CCC) 'requires the differences Xi – Yi to be small' we revised the statement to say that the CCC is close to one only if the differences Xi – Yi are small.
			Table 1	Reviewer: *Specific recommendation: Table 1. Example 3 seems unnecessary as it is well understood that perfect agreement will result in a perfect correlation (r=1). Response: This example was included so that the reader will better understand the difference between absolute agreement (example 3) and the two types of relational agreement (examples 1 and 2)
			Table 3	Reviewer: *Specific recommendation: Table 3. The mean scores for the symptoms would be of interest. Response: Table 3 has now been revised as requested by both Reviewer #1 and #2. This table now contains mean scores (SDs) and the values of all three coefficients of agreement.

Response to Reviewer #2

Page #	Para #	Line #	Table #	Comment
7	3	16-17		Reviewer: The first is discussion comparing the ICC(A,1) and the CCC. The authors state the CCC is easier to calculate (which I agree is true) than the ICC(A,1) so they refer to the ICC(A,1) as the CCC. *Specific recommendation: Are these numerically the same? If so, that should be stated and shown or at least a reference should be provided. If they are not the same, then the sentence should be changed. Response: The two coefficients are identical. This sentence has been changed and a reference has been added.
				Reviewer: The second issue that should be addressed is the discussion of the results.

			Table 3	*Specific recommendation: It would be useful to show the results of all three methods, and then to compare and contrast the differences between them. Response: This information is now included in an expanded Table 3.
7	1	3		Reviewer: *Specific recommendation: One additional minor point - on page 7, line 11, I believe a should equal 4 not 3. Response: The reviewer is correct and this has been corrected.