

# Visually Evaluating the Measurement Comparability Between Paper-Based and Alternate Versions of Administration of the Lung Function Questionnaire

Theresa Gilligan,<sup>1</sup> Lauren Nelson,<sup>1</sup> Anand A. Dalal,<sup>2</sup> Lori McLeod<sup>1</sup>

<sup>1</sup> RTI Health Solutions, Research Triangle Park, NC, United States; <sup>2</sup> GlaxoSmithKline, Durham, NC, United States

## INTRODUCTION

- Intraclass correlation coefficients (ICCs) are often used to assess the concordance between scores on different administration versions of patient-reported outcome measures.
- An ICC provides researchers with a quick and easy way to interpret concordance (one statistic and perhaps a confidence interval [CI]).
- This strength of the ICC—its simplicity—may also be a weakness. Analysts may overlook important details (e.g., biases, outliers) when ICCs are used as the primary method for assessing concordance.
- Bland-Altman plots provide an additional view of the data that highlights biases and outliers.
- Using these methods in tandem provides researchers with a more holistic view of the data and empowers them to make more informed conclusions about the concordance of measures.

## ICCs

- ICCs are often used to assess the concordance between scores on different administrations of the same instrument.
- ICCs should be used when measures share a metric and variance.<sup>1</sup>
- It is generally recommended that ICCs be at least 0.70 for multiple-item scales.<sup>2</sup>

## ICC Models

- Various models can be used to compute ICCs based on:
  - Type of measurement (single or average)
  - Fixed or random effects
  - Presence or absence of interactions
  - Need to assess absolute or consistent agreement.
- The examples displayed in this poster are based on a two-way mixed approach.

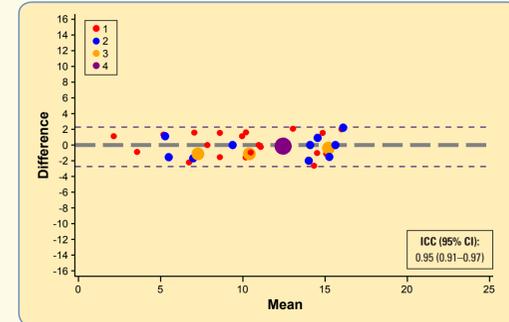
## Bland-Altman Plots

- Bland-Altman plots visually display agreement between two measures.
- The plots display the difference in two administrations (y-axis) against the mean of the two administrations (x-axis).
- The size of the points and corresponding numbers indicate the number of participants represented with identical difference and mean scores.

## Interpreting Bland-Altman Plots

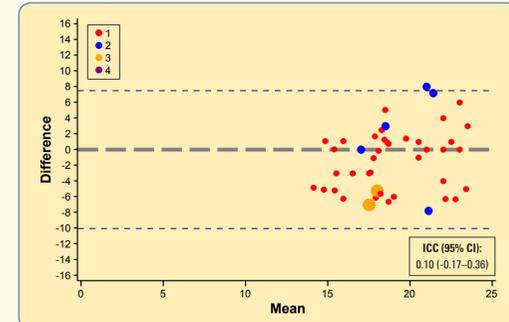
- Measures with greater agreement display points closer to zero across the entire range of mean scores.
- Bias is demonstrated when the points are grouped above or below the zero-difference line (y-axis). Bias indicates that one administration is consistently higher or lower than the other.
- CIs, represented as blue dashed lines, present the level discrepancy (+/- 1.96 x standard deviation) between both measures.
- In an ideal Bland-Altman plot, there is little variation around the zero-difference line (y-axis), indicating that the measures produced very similar results from one measurement to the next (Figure 1).
- The data points are within the Bland-Altman CIs, illustrating that the mean difference between each measurement is within a 95% CI (Figure 1).

Figure 1. Example of an Ideal Bland-Altman Plot



- In a substandard Bland-Altman plot, although the data points are found within a small mean range, they are widely dispersed across the zero-difference line (y-axis), indicating a lot of variation between measurements (Figure 2).

Figure 2. Example of a Substandard Bland-Altman Plot and ICC



## EXAMPLE OF MATCHING ICCS AND BLAND-ALTMAN PLOTS: LUNG FUNCTION QUESTIONNAIRE (LFQ)

### Study Design

- 149 participants who were aged 40 years and older and were self-reported current or former smokers completed the paper-based LFQ.
- Participants were also randomly assigned to complete one of three alternate modes:
  - Web
  - Interview
  - IVRS.

### LFQ

- The LFQ is a five-item questionnaire that was developed using questions from the third National Health and Nutrition Examination Survey (NHANES III).
- The instrument measures patient perception of breathing problems and activity limitation.
- The five items are summed to create a total LFQ score, which can range from 5 to 25.
- Lower scores indicate risk of obstruction.
- For the current study, a total score of 18 or less was indicative of a greater risk for airway obstruction.
- Response category wording was exactly the same for each of the four administration modes.

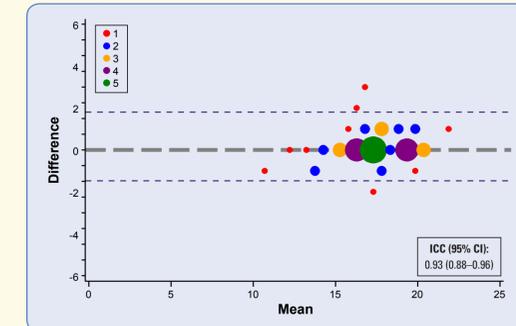
Table 1. LFQ Items

Please think about how you are feeling physically when you are experiencing these symptoms....	Verbal Anchors and Scoring Rubric				
	5	4	3	2	1
1. How often do you cough up mucus?	Never	Rarely	Sometimes	Often	Very often
2. How often does your chest sound noisy (wheezing, whistling, rattling) when you breathe?	Never	Rarely	Sometimes	Often	Very often
3. How often do you experience shortness of breath during physical activity (walking up a flight of stairs or walking up an incline without stopping to rest)?	Never	Rarely	Sometimes	Often	Very often
4. How many years have you smoked?	Never smoked	≤ 10 years	11-20 years	21-30 years	> 30 years
5. What is your age?	< 40 years	40-49 years	50-59 years	60-69 years	≥ 70 years

### LFQ ICC and Bland-Altman Results

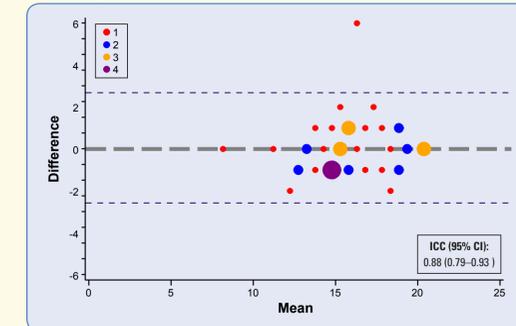
- The ICC associated with the paper-web group was found to be the highest of all the mode comparisons (Figure 3).
- The Bland-Altman plot reflected the strength of the ICC. Most of the points were found on the zero-difference line (y-axis) and were tightly dispersed above and below the line otherwise, indicating low levels of bias (Figure 3).

Figure 3. Paper vs Web



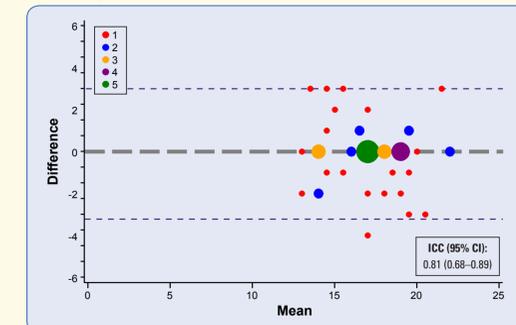
- The reported ICC for the paper-interview group was relatively high, but not as high as the paper-web comparison (Figure 4).
- As indicated by the ICC, we expected to see more variability in the Bland-Altman plot. There were fewer points directly on the zero-difference line (y-axis) and a number of outliers were evident (Figure 4).

Figure 4. Paper Versus Interview



- The paper-IVRS group yielded the lowest ICC in the LFQ study, but the ICC was found to be highly acceptable (Figure 5).
- The Bland-Altman plot reflected the ICC's value in that points were dispersed in a more vertical fashion, indicating more differences between measurements (Figure 5).

Figure 5. Paper Versus IVRS

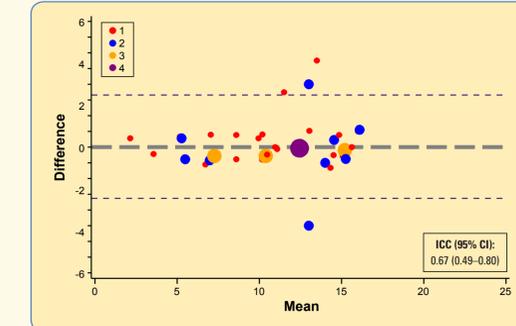


## EXAMPLE OF MISMATCHING ICCS AND BLAND-ALTMAN PLOTS

### Why Should We Look At Both?

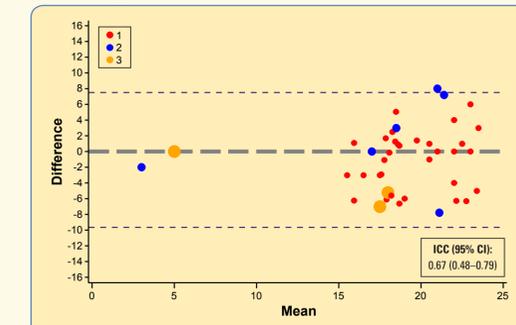
- Figure 6 presents the same plot as the ideal sample (Figure 1), but with a few large differences on the y-axis, indicating a few mismatched measurements.
- A few outliers above and below the confidence intervals dropped the ideal Bland-Altman plot's ICC from .95 to .67.
- By noting these outliers and reviewing the data, it is possible that some of these outliers are erroneous and could be dropped. By looking at the ICC only, we could not draw this conclusion.

Figure 6. Example 1



- Figure 7 presents the same plot as the substandard example (Figure 2), but with a few mean outliers less than 5 on the zero-difference scale. These outliers made a significant difference in the ICC, which is now .67.

Figure 7. Example 2



## CONCLUSIONS

ICCs and Bland-Altman plots complement each other's strengths:

- ICCs provide an efficient and concise estimate to determine the comparability of versions.
- Bland-Altman plots provide a greater level of detail that incorporates a broader view of the analyzed distributions.
- The use of the two methods together provides a more holistic view of concordance.

## REFERENCES

- McGraw KO, Wong SP. Forming inferences about some intraclass correlation coefficients. *Psychol Methods*. 1996;1(1):30-46.
- Nunnally JC, Bernstein IH. *Psychometric theory*. 3rd ed. New York: McGraw-Hill; 1994.

## CONTACT INFORMATION

Theresa Gilligan, MS  
Senior Health Outcomes Analyst  
RTI Health Solutions  
200 Park Offices Drive  
Research Triangle Park, NC 27709  
Phone: +1.919.316.3843  
Fax: +1.919.541.7222  
E-mail: Gilligan@rti.org

Presented at: ISPOR 16th Annual International Meeting  
May 21-25, 2011  
Baltimore, MD, United States