



ELSEVIER

Cracking the Code: Providing Insight Into the Fundamentals  
of Research and Evidence-Based Practice

# A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research



Terry K. Koo, PhD<sup>a,\*</sup>, Mae Y. Li, BPS<sup>b</sup>

<sup>a</sup> Director & Associate Professor, Foot Levelers Biomechanics Research Laboratory, New York Chiropractic College, Seneca Falls, NY

<sup>b</sup> DC Candidate, Foot Levelers Biomechanics Research Laboratory, New York Chiropractic College, Seneca Falls, NY

Received 30 July 2015; received in revised form 3 November 2015; accepted 9 November 2015

## Key indexing terms:

Reliability and validity;  
Research;  
Statistics

## Abstract

**Objective:** Intraclass correlation coefficient (ICC) is a widely used reliability index in test-retest, intrarater, and interrater reliability analyses. This article introduces the basic concept of ICC in the content of reliability analysis.

**Discussion for Researchers:** There are 10 forms of ICCs. Because each form involves distinct assumptions in their calculation and will lead to different interpretations, researchers should explicitly specify the ICC form they used in their calculation. A thorough review of the research design is needed in selecting the appropriate form of ICC to evaluate reliability. The best practice of reporting ICC should include software information, “model,” “type,” and “definition” selections.

**Discussion for Readers:** When coming across an article that includes ICC, readers should first check whether information about the ICC form has been reported and if an appropriate ICC form was used. Based on the 95% confident interval of the ICC estimate, values less than 0.5, between 0.5 and 0.75, between 0.75 and 0.9, and greater than 0.90 are indicative of poor, moderate, good, and excellent reliability, respectively.

**Conclusion:** This article provides a practical guideline for clinical researchers to choose the correct form of ICC and suggests the best practice of reporting ICC parameters in scientific publications. This article also gives readers an appreciation for what to look for when coming across ICC while reading an article.  
© 2016 National University of Health Sciences.

## Introduction

Before any measurement instruments or assessment tools can be used for research or clinical applications, their reliability must be established. *Reliability* is defined as the extent to which measurements can be replicated.<sup>1</sup>

\* Corresponding author: Terry K Koo, PhD, 2360 State Route 89, Seneca Falls, NY, 13148.

E-mail address: tkoo@nycc.edu (T. K. Koo).

In other words, it reflects not only degree of correlation but also agreement between measurements.<sup>2,3</sup> Mathematically, reliability represents a ratio of true variance over true variance plus error variance.<sup>4,5</sup> This concept is illustrated in Table 1. As indicated in the calculation, reliability value ranges between 0 and 1, with values closer to 1 representing stronger reliability. Historically, Pearson correlation coefficient, paired *t* test, and Bland-Altman plot have been used to evaluate reliability.<sup>3,6-8</sup> However, paired *t* test and Bland-Altman plot are methods for analyzing agreement, and Pearson correlation coefficient is only a measure of correlation, and hence, they are nonideal measures of reliability. A more desirable measure of reliability should reflect both degree of correlation and agreement between measurements. Intraclass correlation coefficient (ICC) is such as an index.

Intraclass correlation coefficient was first introduced by Fisher<sup>9</sup> in 1954 as a modification of Pearson correlation coefficient. However, modern ICC is calculated by mean squares (ie, estimates of the population variances based on the variability among a given set of measures) obtained through analysis of variance. Nowadays, ICC has been widely used in conservative care medicine to evaluate interrater, test-retest, and intrarater reliability (see Table 2 for their definitions).<sup>10-17</sup> These evaluations are fundamental to clinical assessment because, without them, we have no confidence in our measurements, nor can we draw any rational conclusions from our measurements.

There are different forms of ICC that can give different results when applied to the same set of data, and the ways for reporting ICC may vary between researchers. Given that different forms of ICC involve distinct assumptions in their calculations and will lead to different interpretations, it is important that researchers are aware of the correct application of each form of ICC, use the appropriate form in their analyses, and accurately report the form they used. The purpose of this article is to provide a practical guideline for clinical researchers to choose the correct form of ICC for their reliability analyses and suggest the best practice of reporting ICC parameters in scientific

**Table 1** Hypothetical Flexion-Extension Range of Motion (ROM) of L4-L5 Measured by Radiograph

Subject	Measured ROM	True ROM	Error
1	28°	28°	0°
2	20°	20°	0°
3	24°	20°	4°
4	18°	22°	-4°
5	26°	22°	4°
6	16°	20°	-4°
Variance	22.4°	9.6°	12.8°

$$\text{Reliability index} = \frac{\text{true variance}}{\text{true variance} + \text{error variance}} = \frac{9.6}{9.6 + 12.8} = 0.43.$$

**Table 2** Definitions of Different Types of Reliability

Types	Definitions
Interrater reliability	It reflects the variation between 2 or more raters who measure the same group of subjects.
Test-retest reliability	It reflects the variation in measurements taken by an instrument on the same subject under the same conditions. It is generally indicative of reliability in situations when raters are not involved or rater effect is neglectable, such as self-report survey instrument.
Intrarater reliability	It reflects the variation of data measured by 1 rater across 2 or more trials.

publications. This article also aims to guide readers to understand the basic concept of ICC so that they can apply it to better interpret the reliability data while reading an article with related topics.

## Discussion for Researchers

### How to Select the Correct ICC Form for Interrater Reliability Studies

McGraw and Wong<sup>18</sup> defined 10 forms of ICC based on the “Model” (1-way random effects, 2-way random effects, or 2-way fixed effects), the “Type” (single rater/measurement or the mean of *k* raters/measurements), and the “Definition” of relationship considered to be important (consistency or absolute agreement). These ICC forms and their formulation are summarized in Table 3.

Selection of the correct ICC form for interrater reliability study can be guided by 4 questions: (1) Do we have the same set of raters for all subjects? (2) Do we have a sample of raters randomly selected from a larger population or a specific sample of raters? (3) Are we interested in the reliability of single rater or the mean value of multiple raters? (4) Do we concern about consistency or agreement? The first 2 questions guide the “Model” selection, question 3 guides the “Type” selection, and the last question guides the “Definition” selection.

#### (A) “Model” Selection

##### One-Way Random-Effects Model

In this model, each subject is rated by a different set of raters who were randomly chosen from a larger population of possible raters. Practically, this model is rarely used in clinical reliability analysis because

Download English Version:

<https://daneshyari.com/en/article/2619939>

Download Persian Version:

<https://daneshyari.com/article/2619939>

[Daneshyari.com](https://daneshyari.com)