

## Reliability Analysis in the Field of Radiology; Common Mistakes

Siamak Sabour

Siamak Sabour, Safety Promotion and Injury Prevention Research Center, Shahid Beheshti University of Medical Sciences, Tehran, I.R. Iran

Siamak Sabour, Department of Clinical Epidemiology, School of Health, Shahid Beheshti University of Medical Sciences, Tehran, I.R. Iran

Correspondence to: Siamak Sabour, Safety Promotion and Injury Prevention Research Center, Department of Clinical Epidemiology, School of Health, Shahid Beheshti University of Medical Sciences, Tehran, I.R. Iran.

Email: [s.sabour@sbmu.ac.ir](mailto:s.sabour@sbmu.ac.ir)

Telephone: +98-21- 22421814

Received: November 3, 2014

Revised: December 16, 2014

Accepted: December 19, 2014

Published online: June 2, 2015

### ABSTRACT

Reliability (precision) is an important methodological issue in the field of radiology which is being assessed by inappropriate tests such as Pearson r, least square and paired t.test. For quantitative variable Intra Class Correlation Coefficient (ICC) and for qualitative variables weighted kappa should be used; however, Bland- Altman as well as coefficient of variance (CV) is also being considered.

© 2015 ACT. All rights reserved.

**Key words:** Reliability; Mistake

Sabour S. Reliability Analysis in the Field of Radiology; Common Mistakes. *International Journal of Radiology* 2015; 1(1): 17-19  
Available from: URL: <http://www.ghrnet.org/index.php/ijr/article/view/1229>

### Advances in Knowledge

1. Reliability (precision) is an important methodological issue in the field of radiology.

2. The reliability is being assessed by inappropriate tests which all of them are among common mistakes and is being published by high impact journals.

3. For reliability analysis, appropriate tests should be applied by clinical researchers.

### Implication for patient care:

Misdiagnosis and mismanagement of the patients in routine clinical care cannot be avoided using inappropriate tests to assess reliability

### RELIABILITY ANALYSIS IN THE FIELD OF RADIOLOGY; COMMON MISTAKES

Reliability (precision) is an important methodological issue in the field of radiology. Reliability (repeatability or reproducibility) is being assessed by different statistical tests such as Pearson r, least square and paired t.test which all of them are among common mistakes in reliability analysis (Figures 1 and 2)<sup>[1]</sup> and is being published by high impact radiology journals<sup>[2-6]</sup>.

Briefly, for quantitative variable Intra Class Correlation Coefficient (ICC) and for qualitative variables weighted kappa should be used; however, Bland- Altman as well as coefficient of variance (CV) is being considered<sup>[7-15]</sup>.

It is important to know that simple kappa has its own limitation too<sup>[1,7-15]</sup>. There is no value of kappa that can be regarded universally as indication good agreement. Two important weaknesses of k value to assess agreement of a qualitative variable are as follow: It depends upon the prevalence in each category which means it can be possible to have different kappa value having the same percentage for both concordant and discordant cells! Table 1 shows that in both (a) and (b) situations the prevalence of concordant cells are 80% and discordant cells are 20%, however, we get different kappa value (0.38 as fair and 0.60 as moderate - good) respectively<sup>[1]</sup>. Kappa value also depends upon the number of categories which means the higher the categories, the lower the amount of kappa value<sup>[7-15]</sup>.

Regarding reliability or agreement, it is crucial to know that an individual based approach instead of group based should be

considered<sup>[2-3]</sup>. The reason is in reliability assessment; we should consider individual results and not global average. Therefore, ICC single measure instead of average measure should be reported to correctly assess the reliability. In other words, possibility of getting exactly the same average of a variable between two methods or observers with no reliability at all is high. The same reasoning is also true for CV<sup>[7-9,14]</sup>.

As a take home message, for reliability analysis, appropriate tests should be applied by radiologists. Otherwise, misdiagnosis and mismanagement of the patients cannot be avoided.

## CONFLICT OF INTERESTS

There are no conflicts of interest with regard to the present study.

## REFERENCES

- Lawrence I, Kuei Lin. A Concordance Correlation Coefficient to Evaluate Reproducibility. *BIOMETRICS*, 1989, March, 45, 255-268
- Roujol S, Weingärtner S, Foppa M, Chow K, Kawaji K, Ngo LH, Kellman P, Manning WJ, Thompson RB, Nezafat R. Accuracy, Precision, and Reproducibility of Four T1 Mapping Sequences: A Head-to-Head Comparison of MOLLI, ShMOLLI, SASHA, and SAPHIRE. *Radiology*. 2014 Sep;272(3):683-9. doi: 10.1148/radiol.14140296. Epub 2014 Apr 4.
- Jain V, Duda J, Avants B, Giannetta M, Xie SX, Roberts T, Detre JA, Hurt H, Wehrli FW, Wang DJ. Longitudinal Reproducibility and Accuracy of Pseudo-Continuous Arterial Spin-labeled Perfusion MR Imaging in Typically Developing Children. *Radiology*, 2012 May; 263(2):527-36.
- Albarakati SF, Kula KS, Ghoneima AA, The reliability and reproducibility of cephalometric measurements: a comparison of conventional and digital methods. *Dentomaxillofac Radiol*, 2012 Jan; 41(1):11-7.
- Ling LF, Obuchowski NA, Rodriguez L, Popovic Z, Kwon D, Marwick TH. Accuracy and Interobserver Concordance of Echocardiographic Assessment of Right Ventricular Size and Systolic Function: A Quality Control Exercise. *J Am Soc Echocardiogr*, 2012 Apr 26
- Maislin G, Ahmed MM, Gooneratne N, Thorne-Fitzgerald M, Kim C, Teff K, Arnardottir ES, Benediktsdottir B, Einarsdottir H, Juliusson S, Pack AI, Gislason T, Schwab RJ. Single Slice vs. Volumetric MR Assessment of Visceral Adipose Tissue: Reliability and Validity Among the Overweight and Obese. *Obesity (Silver Spring)*, 2012 Mar 7. doi: 10.1038/oby.2012.53.
- Jeckel J.F, Katz D.L, Elmore J.G, Wild D.M.G, *Epidemiology, Biostatistics and Preventive Medicine*, 3rd edition. 2007, SAUNDERS, Elsevier, Philadelphia, PA, United State
- Kenneth J. Rothman, Sander Greenland, Timothy L. Lash. *Modern Epidemiology*, 4th edition. 2010. Lippincott Williams & Wilkins, Baltimore, United States
- Szklo M, Nieto F.J, *Epidemiology beyond the basics*, 2 nd edition, 2007, Jones and Bartlett Publisher, Manhattan, new York, United State
- Sabour S, Kermani H, Accuracy of linear intraoral measurements using cone beam CT and multidetector CT: methodological mistake. *Dentomaxillofac Radiol*. 2013;42(4):20130048. doi: 10.1259/dmfr.20130048. Epub 2013 Feb 18. No abstract available.
- Sabour S, A quantitative assessment of the accuracy and reliability of O-arm images for deep brain stimulation surgery. *Neurosurgery*. 2013 Apr;72(4):E696.
- Sabour S. Reproducibility of the external surface position in left-breast DIBH radiotherapy with spirometer-based monitoring: methodological mistake. *J Appl Clin Med Phys*. 2014 Jul 8;15(4):4909. doi: 10.1120/jacmp.v15i4.4909.
- Sabour S. Methodologic concerns in reliability of noncalcified

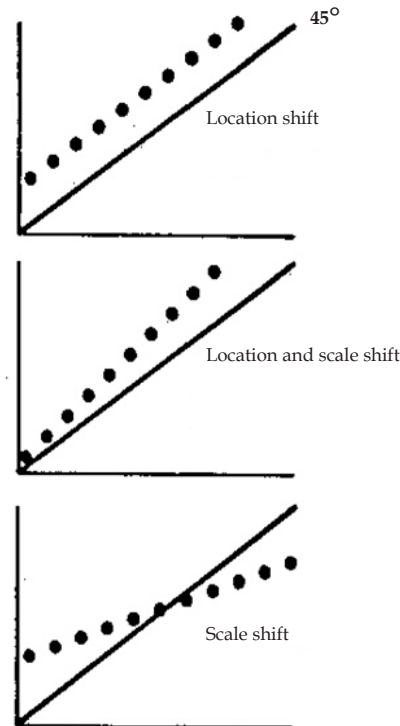


Figure 1 Cases when Pearson correlation coefficient fails to detect non reproducibility.

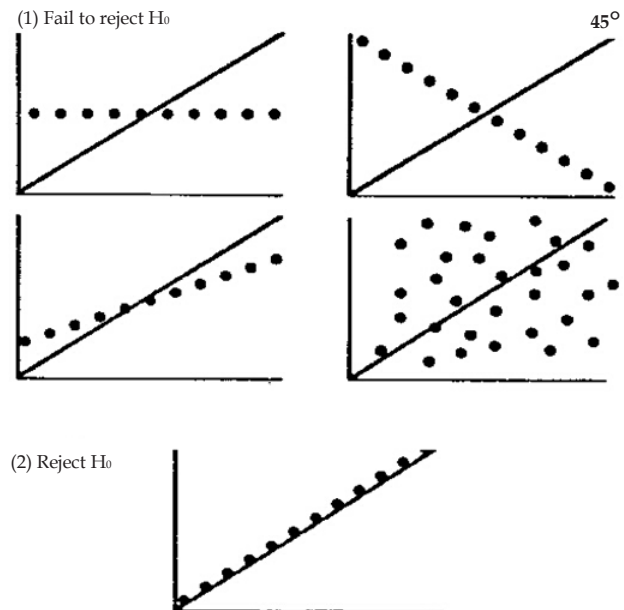


Figure 2 Cases when paired t-test can be misleading. H<sub>0</sub>: Means are equal.

Table 1 Comparison of two observers' diagnosis with different prevalence in the two categories.

		Observer 1		
(a)		Positive	Negative	Total
Observer 2	Positive	70	10	80
	Negative	10	10	20
	Total	80	20	100
K=0.38				
		Observer 1		
(b)		Positive	Negative	Total
Observer 2	Positive	40	10	50
	Negative	10	40	50
	Total	50	50	100
K=0.60				

- coronary artery plaque burden quantification. *AJR Am J Roentgenol.* 2014 Sep;203(3):W343. doi: 10.2214/AJR.14.12649.
14. Sabour S. Validity and reliability of the <sup>13</sup>C-methionine breath test for the detection of moderate hyperhomocysteinemia in Mexican adults; statistical issues in validity and reliability analysis. *Clin Chem Lab Med.* 2014 Jun 14. pii: /j/cclm.ahead-of-print/cclm-2014-0453/cclm-2014-0453.xml. doi: 10.1515/cclm-2014-0453. [Epub ahead of print]
  15. Sabour S, Ghassemi F. Reliability of on-call radiology residents' interpretation of 64-slice CT pulmonary angiography for the detection of pulmonary embolism: methodological error. *Acta Radiol.* 2014 May;55(4):427. doi: 10.1177/0284185114521413.

**Peer reviewers:** Janney Sun, Professor, Editor-In-Chief, Int Journal of Radiology, Unit A1, 7/F, Cheuk Nang Plaza, 250 Hennessy Road, Wanchai, Hong Kong.