

Other

2010-01-01

The Statistical Interpretation of the Coefficient of Repeatability

James Loughman

Technological University Dublin, james.loughman@tudublin.ie

Follow this and additional works at: <https://arrow.tudublin.ie/otpomoth>



Part of the [Optometry Commons](#)

Recommended Citation

Loughman, James, "The Statistical Interpretation of the Coefficient of Repeatability" (2010). *Other*. 1.
<https://arrow.tudublin.ie/otpomoth/1>

This Other is brought to you for free and open access by ARROW@TU Dublin. It has been accepted for inclusion in Other by an authorized administrator of ARROW@TU Dublin. For more information, please contact arrow.admin@tudublin.ie, aisling.coyne@tudublin.ie, vera.kilshaw@tudublin.ie.

1. The Statistical Interpretation of the Coefficient of Repeatability

(Loughman, James)

Bartlett et al. have reported on the clinical viability of the MPS 9000, a heterochromatic flicker photometry (HFP) device used to estimate macular pigment optical density (MPOD).[1] Coefficient of repeatability (CoR) values were used to determine the test-retest repeatability and reproducibility of the MPS 9000. While the observed CoR of 0.33 is indeed high, and, superficially at least, raises concern regarding the clinical application and interpretation of the device, I would suggest that the author's interpretation of their data, and resultant conclusion in this paper, that a change in MPOD of less than 0.33 "is very likely to be due to measurement noise", is both misleading and inaccurate.

The reported CoR (ranging from 0.25 to 0.33) is in fact a measure of the 95% limits of agreement (LoA), as proposed originally by Bland and Altman, and is calculated as the mean \pm 1.96SD.[2] By definition, this value provides an interval, within which 95% of test-retest measurement differences lie, in this case 0.33.[2-4]

On the basis of this reported coefficient of repeatability, the authors suggest that any change in repeated measures of MPOD of less than 0.33 on the MPS 9000, should be interpreted as measurement noise, and could not be assumed to be of clinical importance.[1] In other words, the authors interpret their coefficient of repeatability values as an indicator of the amount of change that can occur between readings and still be classed as measurement noise. Such a conclusion would suggest a fundamental lack of understanding of this useful statistical tool on the Author's behalf. Confining the analysis to increases in MPOD (as would be expected clinically in response to dietary modification or supplementation), the simple interpretation of a CoR value of 0.33 is that, in the test-retest data, the observed increase in MPOD was less than 0.33 for 97.5% of subjects, or conversely, that the probability of detecting a test-retest increase in MPOD greater than 0.33 in the test population is only 2.5%.

Expanding on this statistical interpretation further, it can be seen that if the CoR value is halved to 0.17 (mean + 1SD), 84% of subjects would be expected to exhibit retest increases less than this 0.17 value. If, for example, during routine clinical practice, a retest increase in MPOD of 0.17 was noted for a particular patient, on the basis of the Bartlett et al. results,[1] one could interpret that the probability of such a change being due to measurement noise is as little as 16%, and for an increase of 0.33, the probability is as little as 2.5%. In other words, on the basis of probability, such differences would more than likely represent a genuine change in MPOD rather than measurement noise as is suggested by Bartlett et al.[1]

Extracting the data from the Bartlett et al. paper, [1] it can be observed that the test - retest difference is < 0.1 in close to two- thirds of their subject data. The data also reveals a significant influence of two outliers on the magnitude of the observed CoR. For one subject, the difference in MPOD between visits is an incredible 0.69, and for another, the difference is greater than 0.4 in the opposite direction. Simple exclusion of these outliers dramatically improves the CoR to ~ 0.23 . Similar interpretations have been adopted by the same group in a number of recent papers, and ultimately, the general validity of the HFP technique, and its applicability to clinical practice have thereby been called into question by the group.[5] Given that (a) the results here from just two subjects are observed to have a significant and adverse effect on the reported CoR, (b) the test-retest values are quite repeatable for a very high percentage of subjects, and (c) the authors repeatedly demonstrate a lack of understanding of the appropriate statistical interpretation of their data, their broad conclusions would, therefore, seem neither reasonable nor sustainable.

The suggestion that "any change less than 0.33 units should not be considered clinically significant as it is very likely to be due to measurement noise" is simply incorrect. A more appropriate conclusion to this paper might have been to suggest that, for clinical practice, a number of measurements of MPOD might be used to provide an average MPOD value, to thereby ensure that the value obtained is maximally robust, and that cases of obviously poor performance on the test can be identified.

References

1. Bartlett H, Stainer L, Singh S et al. Clinical evaluation of the MPS 9000 Macular Pigment Screener. *Br J Ophthalmol* 2010;94:753-6.
2. Bland J, Altman D. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1986;1:307-10.
3. Bland J, Altman D. Measuring agreement in method comparison studies. *Stat Methods Med Res* 1999;8:135-60.
4. Atkinson G, Nevill AM. Statistical methods for assessing measurement error (reliability) in variables relevant to sports medicine. *Sports Med* 1998;26(4):217-38.
5. Bartlett H, Howels O, Eperjesi F. The role of macular pigment in clinical practice: a review. *Clin Exp Optom* 2010;93(5):300-8

Conflict of Interest:

None declared

[... less](#)

[Submit response](#)

Published 22 February 2011

