JP Journal of Biostatistics
© 2014 Pushpa Publishing House, Allahabad, India
Published Online: March 2015
Available online at http://www.pphmj.com/journals/jpb.htm
Volume 11, Number 2, 2014, Pages 135-141

J-MEASURES FOR ASSESSING AGREEMENT BETWEEN TWO METHODS OF CLINICAL MEASUREMENT

Farzad Noubary[1] **and Reza Noubary**[2]

[1]The Institute for Clinical Research and Health Policy Studies
Tufts Medical Center
U. S. A.

[1]Tufts Clinical and Translational Science Institute
Tufts University
Boston, MA, U. S. A.

[2]Department of Mathematics, Computer Science and Statistics
Bloomsburg University
Bloomsburg, PA, U. S. A.

Abstract

A statistical method is proposed for assessing agreement between two clinical methods of measuring an unknown quantity. Its value is a function of the difference between measurements of the same sample by the two methods and their arithmetic and geometric means. The proposed method is motivated by the *J*-divergence rate and can be viewed as a measure of the distance between two distributions defined by measurements by the two methods. The paper concludes with a discussion regarding the simultaneous use of this and Bland-Altman methods.

Received: November 8, 2014; Accepted: December 1, 2014

Keywords and phrases: measure of agreement, geometric mean, *J*-divergent, Bland-Altman, information.

Introduction

Medical researchers often need to compare two methods of measuring an unknown quantity. A widely used method developed by Bland and Altman plots the difference of measurements against their arithmetic mean (Altman and Bland [1, 3]). Liao and Capen [4] have tried to improve Bland-Altman method by defining an agreement interval for each individual pair and assessing the overall concordance. In what follows, we suggest a different approach and point out its strength.

Suppose that w_1 and w_2 are measurements on the same subject using the reference standard and a novel method, respectively. Here both the arithmetic mean $(w_1 + w_2)/2$ and the geometric mean $\sqrt{w_1 w_2}$ provide estimates of the true but unknown quantity of interest. The following facts provide insight regarding these estimates:

Fact 1.

$$(w_1 + w_2)/2 \geq \sqrt{w_1 w_2}.$$

Fact 2. The arithmetic mean $(w_1 + w_2)/2,$ the geometric mean $\sqrt{w_1 w_2},$ and difference $d = w_2 - w_1$ have the following relationship:

$$(d/2)^2 = ((w_2 - w_1)/2)^2 = ((w_1 + w_2)/2)^2 - (\sqrt{w_1 w_2})^2.$$

Moreover,

$$\begin{aligned}(d/2)^2 &= ((w_1 + w_2)/2 + \sqrt{w_1 w_2})((w_1 + w_2)/2 - \sqrt{w_1 w_2}) \\ &= 1/4(\sqrt{w_1} + \sqrt{w_2})^2(\sqrt{w_2} - \sqrt{w_1})^2.\end{aligned}$$

These facts shed light on the appropriateness of plotting the difference against the mean of two measurements. Using these facts, we may consider working with the difference and geometric mean, or the arithmetic mean and geometric mean, noting that d is a function of the sum and difference of the arithmetic mean and geometric mean. Additionally, if we consider a right triangle with legs $\sqrt{w_1 w_2}$ and $(w_2 - w_1)/2$ (if $w_2 > w_1$) or $(w_1 - w_2)/2$ (if $w_1 > w_2$) and let θ denote the angle opposite the shorter leg, $(w_2 - w_1)/2,$

then we have $\text{Sin}\,\theta = (w_2 - w_1)/(w_2 + w_1)$, $\text{Cos}\,\theta = 2\sqrt{w_1 w_2}/(w_2 + w_1)$, and $\text{Tan}\,\theta = (w_2 - w_1)/2\sqrt{w_1 w_2}$. Thus, the comparison can also be made using these functions or just θ itself. The closer the θ is to 0 the better the agreement. We can also develop a method in line with Liao and Capen [4] using θ values in place of their differences. This would improve their method since statistics based on θ are scale invariant.

A Measure of Agreement

The measure (*J*-measure) for assessing the agreement given below is motivated by the above facts and the *J*-divergence rate. We suggest using the following:

$$J = \frac{w_1}{w_2} + \frac{w_2}{w_1} - 2 = \frac{(w_2 - w_1)^2}{w_1 w_2} = \left[\frac{(w_2 - w_1)}{\sqrt{w_1 w_2}}\right]^2$$

$$= 4\left[\left[\frac{(w_1 + w_2)/2}{\sqrt{w_1 w_2}}\right]^2 - 1\right] = 4\text{Tan}^2\theta \geq 0.$$

We think this is a reasonable approach as the agreement problem can be viewed as a discrimination problem. Note that the *J*-measure is a function of the arithmetic mean, geometric mean, their ratio, and the normalized differences. For example, for a fixed value of the geometric mean, it increases directly with the differences (*d*). Thus, for data with fixed geometric mean, plotting *J* against the arithmetic mean produces the same information as a Bland-Altman plot.

The use of *J*-measure has similarity with the Bland-Altman method in that their method involves the difference and arithmetic mean, whereas the *J*-measure involves the difference and geometric mean. Rather than plotting *d* against the arithmetic mean, we can plot the ratio of the two measurements against the geometric mean. The closer the ratio is to 1, the better the agreement. Clearly, plotting the latter in log-scale is the same as Bland-Altman method. When applying their method to a set of real data, Bland and Altman [3] have noted that results related to differences between log percentages are not easy to interpret. They have suggested back-transforming (antilog) the results to get values relating the ratios of measurements by the two methods. We can make the transformation process more transparent by working directly with the ratios. Instead of taking logs and calculating differences, we can simply calculate the ratio of the two values for each subject and calculate limits of agreement based on the mean and standard deviation of these or, as is suggested here, work with the *J*-measure.

We end this section by noting that, using the above relationships, we could estimate the proportional bias using either of the following:

$$1/N\left(\sum_1^N (w_{2i} - w_{1i})/((w_{1i} + w_{2i})/2)\right)$$

$$1/N\left(\sum_1^N (w_{2i} - w_{1i})/(\sqrt{w_{1i}w_{2i}})\right).$$

Why *J*-measure

The scale-dependent limits of agreement, $\bar{d} \pm 1.96 s_d$ (where $\bar{d}$ is the mean of the differences and s_d is their standard deviation), suggested by Bland and Altman assigns equal weight to all measured differences. However, it may be reasonable to think that a difference of $d = 20$ resulting from the measurements, $w_1 = 600$, $w_2 = 120$ is a "stronger" indicator of disagreement between the two methods than the same difference resulting from the measurements $w_1 = 600$, $w_2 = 620$. Transformations such as square-root or logarithm can help this problem when data values are positive. For the above measurements, the difference of 20 translates to square-root differences of 0.954 and 0.405 and log-differences of 0.079 and 0.014, respectively. In fact, since

$$w_2 - w_1 = (\sqrt{w_2} + \sqrt{w_1})(\sqrt{w_2} - \sqrt{w_1}),$$

we see that *d* weights the transformed data by the sum of their sizes.

The *J*-measure accounts for the above in a reasonable fashion. For the measurements quoted above, the *J* values are 0.033 and 0.001, respectively.

Here 95% limits of agreement using the *J*-measure is

$$\bar{J} \pm 1.96 s_j.$$

A 95% confidence interval for limits of agreement can also be constructed following the same derivation as the Bland-Altman method. We note that if measurements were proportional, that is $w_2 = kw_1$, then $J = k + \frac{1}{k} - 2$ for all measurements and no limit of agreement can be constructed.

Values of this measure provide us with quantities indicating the contribution of pairs of measurements to the *J*-measure. In fact, if we consider the problem as a discrimination issue, then we can identify which specific pair or which group of pairs contributes most to the disagreement.

Threshold Agreement Using *J*-measure

Though researchers are typically interested in assessing agreement throughout the entire range of measurements, one may be particularly interested in agreement relative to a clinical threshold of interest. For these cases, we can just consider *J* values in the vicinity of the threshold. In such a scenario, the *J*-measure could prove useful, noting that for a fixed difference between measurements on the same subject, *J* is larger for smaller values of *w* than for larger values. That is, it gives a bigger weight to a fixed difference occurring in lower ranges than the high ranges. Using this, we can either subtract the threshold value from both w_1 and w_2 and move the origin to the threshold of interest or just consider *J* values in vicinity of the threshold.

J-measure and Information

Recall that if $w_2 = kw_1$, then *J*-measure has a fixed value $J = k + \frac{1}{k} - 2$ independent of the size of the measurements. This is a useful special case since $k > 1$ can be related to the situation where the novel method may be more variable than the reference standard. Note that the larger the w_1, the bigger the difference d. This case is also related to the situation where one may approach the problem using the so-called Lehman alternatives described below.

Suppose that w_1 and w_2 are random variables representing measurements from the reference and novel technologies with cumulative distribution functions, $F_{w1}(x)$ and $F_{w2}(x)$, respectively. Since measurements are taken on the same patient, it is reasonable to assume that $F_{w2}(x) = F_{w1}^k(x)$, where k is a positive constant. For this case, we may use an information theoretic concept and define a divergence measure between $F_{w1}(x)$ and $F_{w1}^k(x)$. Let $f_{w2}(x)$ and $f_{w1}(x)$ denote the probability density functions of w_2 and w_1, respectively, and consider the expressions

$$I_1 = \int_{-\infty}^{\infty} f_{w1}(x) \log(f_{w1}(x)/f_{w2}(x))dx,$$

$$I_2 = \int_{-\infty}^{\infty} f_{w2}(x) \log\left(\frac{f_{w2}(x)}{f_{w1}(x)}\right)dx.$$

These are Kullback-Leibler discrimination information rates for discriminating in favor of $f_{w1}(x)$ over $f_{w2}(x)$ and $f_{w2}(x)$ over $f_{w1}(x)$, respectively. Their sum $J = I_1 + I_2$ is the J-divergence rate and measures the degree of separation between $F_{w1}(x)$ and $F_{w2}(x)$. For distributions $F_{w1}(x)$ and $F_{w1}^k(x)$, the divergence rate is given by

$$J = k + \frac{1}{k} - 2.$$

This is a monotonic increasing function of k with a minimum of zero occurring at $k = 1$. Note that for this situation

$$\rho = P(w_2 > w_1) = \int_0^{\infty} F_{w1}(x)dF_{w2}(x) = k/(1+k)$$

is a measure for bias $(\text{no bias when } \rho = 1/2)$. Also, since $\rho = \frac{k}{(1+k)}$, we have

$$J = (1 - 2\rho)^2/\rho(1 - \rho).$$

This function attains its minimum for $\rho = 1/2,$ and its maximum at $\rho = 1$ and is monotonic increasing in $1/2 \leq \rho \leq 1.$ This shows that the agreement between two technologies can also be defined on the basis of divergent rate between their distributions.

Concluding Remarks

Suppose that $w_2 - w_1 = d$ is constant for all pairs indicating a fixed bias. Using the method developed by Bland-Altman no limits of agreement can be constructed for this case as $\bar{d} = d$ with no variability. Similarly when $w_2 = kw_1$ (proportional bias), the *J*-measure has a fixed value $J = k + \frac{1}{k} - 2$ independent of the size of the measurements and no limits of agreement can be constructed for the *J*-measure. Thus, for cases with systematic bias, no limits of agreement can be constructed. Considering these, it may be beneficial to use Bland-Altman and *J*-measure methods simultaneously. This is particularly useful when measurements come from normal distribution, as for this case, $w_2 = aw_1 + b.$ If a or b or both show significant changes over the range of possible values, then we may construct several limits of agreement and utilize results of piecewise regression.

References

[1] D. G. Altman and J. M. Bland, Measurement in medicine: the analysis of method comparison studies, The Statistician 32 (1983), 307-317.

[2] J. M. Bland and D. G. Altman, Statistical methods for assessing agreement between two methods of clinical measurement, Lancet 2 (1986), 307-310.

[3] J. M. Bland and D. G. Altman, Measuring agreement in method comparison studies, Stat. Methods Med. Res. 8 (1999), 135-160.

[4] J. Z. Liao and R. Capen, An improved Bland-Altman method for concordance assessment, Int. J. Biostat. 7(1) (2011), 1-17.