

0.1 Information Approach

PURPOSE: Disagreement on the interpretation of diagnostic tests and clinical decisions remains an important problem in medicine. As no strategy to assess agreement seems to be fail-safe to compare the degree of agreement, or disagreement,

0.1.1 Example: Systolic Blood Pressure

Bland and Altman (19) present the example of measurements of systolic blood pressure of 85 individuals, by two observers (observer J and observer R) with sphygmomanometer, and one other measurement, by a semiautomatic device (device S). Luiz et al. (16) re-analyze the data and also observe, with a graphical approach, a greater agreement between the two observers than between the observers and the semiautomatic device. Using our information-based measure of disagreement; we also obtained a significantly more disagreement between each observer and the semiautomatic device than between the two observers (Table 1).

0.1.2 Discussion

- We can look at disagreement between observers as the distance between their ratings, so the metric properties are important. Moreover, the proposed measure of disagreement is scale-invariant, i.e., the degree of disagreement between two observers should be the same if the measurements are analyzed in kilograms or in grams, for example.
- Differential weighting is another property of the proposed information-based measure of disagreement: each comparison between two ratings is divided by a normalizing factor, depending on each pair of ratings alone, before summing. Therefore, the information-based measure of disagreement is appropriate for ratio scale measurements (with a natural 0) and it is not appropriate for interval scale measurements (without a natural 0).
- For example, outside air temperature in Celsius (or Fahrenheit) scale does not have a natural 0. The 0 is arbitrary and it does not make sense to say that 20 is twice as hot as 10. Outside air temperature in Celsius (or Fahrenheit) scale is an interval scale. On the other hand, height has a natural 0 meaning: the absence of height. Therefore, it makes sense to say that 80 inches is twice as large as 40 inches. Height is a ratio scale.
- Suppose the heights of a sample of subjects measured independently by two different observers. A difference between the two observers of 1 inch in a child subject represents a worse observers' error than a disagreement between observers of 1 inch in an adult subject.
- Due to differential weighting property of the information-based measure of disagreement, a difference between the observers of one inch in a child

in fact weights less to the estimate of information-based measure of disagreement between observers than a difference between the observers of 1 inch in an adult.

- The usual approaches used to evaluate agreement have the limitation of the comparability of populations. In fact, ICC depends on the variance of the trait in the population; although this characteristic can be considered an advantage it does not permit one to compare the degree of agreement across different populations. Also the interpretation of the limits of agreement depends on what can be considered clinically relevant or not, which could be subjective and different from reader to reader.
- The comparison of the degree of agreement in different populations is not straightforward. Other approaches 16 and 17 to assess observer agreement have been proposed, however the comparability of populations is still not easy with these approaches.
- The proposed information-based measure of disagreement, used as a complement to current approaches for evaluating agreement, can be useful to compare the degree of disagreement among different populations with different characteristics, namely with different variances.
- Moreover, we believe that information theory can make an important contribution to the relevant problem of measuring agreement in medical research, providing not only better quantification but also better understanding of the complexity of the underlying problems related to the measurement of disagreement.