***How Bland and Altman have extended the Difference Plot (with Limits of agreement) to the case of replicate measurements.***

When replicate measurements are taken with each method on each item, more complex methodologies are required. Bland & Altman(1999) describe a number of approaches to adopt in such a case. The first approach is to take the average measurement over the replicates measurements on each subject by each method.

The second approach is to simply treat these replicate measurements as independent measurements in their own right, and use the pre-existing Bland-Altman approach for single measurements.

The proposed approach shares the same overall structure as their earlier work. However complex calculations are now required to compute the variance of case-wise differences. This is compounded by the fact that different approaches are required depending on whether or not there are equal or unequal numbers of replicates.

***What have they missed? Carstensen et al discuss in depth the shortfalls in this approach.***

Carstensen et al (2008) address the flaws in both approaches, demonstrating how the first approach would lead to an underestimation of the variance of case-wise differences. The second case would also lead to incorrect estimates, although it is pointed out that the second approach is much better than the first.

***What are Tolerance Intervals?***

A tolerance interval is a statistical interval within which, with some confidence level, a specified proportion of a population falls.

The ***Engineering Statistics Handbook***describes the difference: *Confidence limits are limits within which we expect a given population parameter, such as the mean, to lie. Statistical tolerance limits are limits within which we expect a stated proportion of the population to lie.*

It is useful to make the distinction between tolerance intervals and confidence intervals clear. The confidence interval describes a single-valued population parameter, commonly the mean, with a specified confidence level. The tolerance interval, on the other hand, describes the range of data values that includes a specific proportion of the population.

As discussed in *Vardeman (1992),* the tolerance interval is not as widely used as the confidence interval and prediction interval, largely because of the emphasis placed on these in undergraduate teaching. Furthermore, *Vardeman(1992)* argues this lack of awareness can lead to misuse of confidence intervals where other types of intervals are more appropriate.

Curiously Carstensen et al (2008) describe the Limits of agreement as a prediction interval, although stating that it is formulated in correctly for that purpose.

***Why Tolerance Intervals are appropriate?***

It is clear from the definition of Tolerance intervals that they function precisely as Bland-Altman intend.

***Total Deviation Index and Coverage Probability***

The Coverage Probability describes the proportion captured within a pre-specified boundary of the absolute paired-measurement differences from two methods of measurement, i.e., the value of ***k*** such that ***P(|D| <k) = pk.***

***Bayesian Approach:***

***consistent with philosophy of hypothesis testing free stats, but it is unclear what Bland and Altman had in mind.***

*Schluter(2009)* develops a multivariate hierarchical Bayesian approach to Bland-Altman’s methodology, presenting two methods of analysis that complement pre-existing literature.