Recommendations for assessing commutability part 2: using the difference in bias between a reference material and clinical samples

Supplemental File: Explanations to the supplemental file spreadsheet Commutability Example Calculations

1. General

Refer to the Supplemental data file Excel workbook "Commutability Example Calculations" for the data and calculations for this example.

This example illustrates application of statistical methods for assessment of commutability using the difference in bias approach. The example describes a case with no trends in the difference plot and should not be used as a template for all applications of this statistical approach for commutability assessment. Rather the spreadsheet is a guide to how to apply this statistical approach. In more complicated situations, the analyst has to make decisions at several steps about how to proceed depending on assessment of the experimental data. All statistical analysis is an interactive process in which one has to inspect the data and decide whether requirements are satisfied for using a statistical method and which application of the method is suitable for the characteristics of the data and the experimental design.

When adapting the spreadsheet for other data, the cells can be adjusted as needed for the amount of data and the equations in cells should be examined to ensure that relevant data are included.

Abbreviations:

C, commutability criterion

CS, clinical sample

RM, reference material

Statistical symbols are in the primary paper for which this is a supplemental file.

2. Sequence of measurements (Sheet Allocation)

In this sheet the sequence of measurements for the 50 clinical samples and 5 reference materials is stated. Each sample is measured in triplicate as three sequential measurements and then the next sample in the sequence is measured in triplicate.

3. **Raw data** (Sheet *Raw_data*)

In this sheet the raw data are presented. We have triplicate measurements of the CSs and triplicate measurements for the RMs that were intermixed in 5 positions among the CSs. For the RMs the ln-transformations are also presented. Replicate 3 for CS 6 (yellow highlight) is to be excluded as an obvious outlier.

4. **Examination of precision and differences for the CSs** (Sheets *CS_Trans and CS_Trans(2)*)

Precision profiles and difference plots are presented both for concentration and ln(concentration) in sheet *CS_Trans*. On the x-axis we always have concentration, for MP x or MP y respectively for precision profiles, and the mean concentration for MP x and MP y for difference plots. It is acceptable to have a log-scale on the x-axis due to the magnitude of the concentration interval.

The precision profiles indicate one outlier for MP x. Examination of the replicates showed that for CS6 one value, 75.4 (yellow highlight), was considerably different from the other two replicates, 7.0 and 6.9 and thus the value 75.4 was deleted. Note that if one value of a triplicate is deleted, the corresponding cell in the ln-transformation table must be emptied (deleted).

On the difference plot for ln(concentration) it is obvious that the difference for CS21 is considerably greater than all others and CS21 (values 7.4-7.7 for MP x and 31.8-32.7 for MP y; yellow highlight) was considered an outlier and deleted for data analysis.

After deleting one replicate for CS6 and all data for CS21 the analysis is presented in sheet *CS_Trans(2)*. The outlying results in the concentration interval 19 -20 have now disappeared from the difference plots.

NOTE 1 If we look at a difference plot for CSs and find that the points scatter around a constant value or a continuous function without outliers it is obvious that a commutable RM should be close to the center of the scatter band. If there are one or more outliers or two groups of CS points separated from each other it will not be possible to find a RM that will be close to the center of the scatter band for all CSs. The best we can do is to find a RM that is close to the majority of the CSs or try to identify different populations of CSs and assess commutability of a RM separately for each of them.

NOTE 2 It is not realistic to state in advance what is an acceptable proportion of outliers to remove from analysis of the data. An informed judgment needs to be made based on the intended use of a RM. It is important to remove outliers that represent unusual performance differences between the two measurement procedures being examined so the statistical assessment will be appropriate. A table of all outlier samples should be maintained and examined for patterns that may indicate limitations in a measurement procedure or identify measurement procedures for which a RM may not be suitable for use. One approach to decide whether to remove outliers is to

perform the statistical analysis with and without questionable outliers and determine if the outliers change the decision regarding commutability of a RM.

We can now examine whether concentration or ln(concentration) is most appropriate for the statistical analysis, that is whether the SD in the precision profiles and the width of the scatter in each difference plot is approximately constant over the concentration. It is obvious that ln(concentration) is to be preferred. Thus, ln-transformed data in *CS_Trans(2)* are used for the statistical analysis in sheet *CS_Diff*.

5. Investigation of trends and sample specific differences for CS (Sheet *CS_Diff*)

If there is a trend in a difference plot, the SD estimated from the successive differences, s_{MSSD}, should be less than the SD of the differences, s_B. The test statistic is $Q = (s_{MSSD}/s_B)^2$. Critical limits are given in the columns I and J for n > 21. For $n \leq 21$ critical limits are found in sheet *Table Q*. In this example, a significant trend is identified because the value for Q is less than the critical value. An inspection of the difference plot shows, however, that the three largest differences are consecutive. As the three largest differences are not in the ends of the concentration interval it seems probable that they happened to be consecutive just by chance. When large or small values happen to be consecutive s_B is not changed but s_{MSSD} will be smaller. If these three CSs are excluded we have the results in sheet *CS_Diff(2)*.

NOTE If rows are deleted the formulas in columns H to J must be corrected, for instance by copying downwards from the first row.

In *CS_Diff(2*), with the three samples removed, there is no longer a trend because Q is now greater than the critical value and the question is whether we should consider the three largest differences as outliers or not. They are not obvious outliers and do not deviate more from the mean than the smallest difference. Thus, it seems to be most reasonable not to remove them and to perform the statistical analysis with all CSs in sheet *CS_Diff*. The bias is, however, considered as constant with no trend based on the assessment in sheet *CS_Diff(2)*, and the bias is estimated by B_{CS} (cell G63 in sheet *CS_Diff*).

NOTE *CS_Diff(2)* is only used to check whether the significant trend can be explained by three CSs with the largest differences.

NOTE When there is no trend s_B may be used instead of s_{MSSD} for estimation of the SD of the sample specific differences, s_d. However, we recommend always using the estimate based on s_{MSSD} because it is more robust to the possibility of a minor but insignificant trend.

NOTE If there had been a trend in the example, a bias function would be estimated by moving averages of q (even) consecutive differences B_i. The bias at the concentration of a RM would be estimated by the moving average with $q/2$ CSs on each side of the value for the RM. In sheet *CS_Diff q* is 12 (a suitable minimum value according to the primary paper for which this is a supplemental file) and the average of the first 12

differences, 0.17, is assigned to the concentration of the 7th CS, CS34, in the sequence. The average of differences 2 to 13 is assigned to the 8th CS, CS28, and so on. For RM1 the mean concentration of both MSs [(x+y)/2 in sheet *Raw_data* cell L55] is 16,3. The first CS with a concentration above this value is CS38 and the CS bias at the concentration of RM1 is estimated as 0.13 (cell I32 in sheet *CS_Diff*).

NOTE Plots of moving averages often erroneously give the impression of a trend because the consecutive averages are not independent (q-1 individual values are the same).

6. Investigation of position effects for RM (Sheet *RM_ANOVA*)

For each RM we have five positions distributed throughout the measurement run with three replicates in each position. First the mean and standard deviation for each position are calculated. The standard deviation of the position means is denoted $s_{Pos\text{-}mean}$ and the pooled standard deviation of the standard deviations within positions is denoted s_e. To test the hypothesis of no position effects the test statistic is

$$F = \frac{3s_{Pos-mean}^2}{s_e^2}$$

The 5% and 1% significance limits for F are given in the sheet using 4 degrees of freedom for the variance between positions (5-1) and 10 degrees of freedom for the variance within positions, (3-1) degrees of freedom in each position multiplied by 5 positions.

The standard deviation of the position effects is estimated as

$$s_{Pos} = \sqrt{s_{Pos-mean}^2 - \frac{s_e^2}{3}}$$

When F<1 (the variance under the root sign is negative) the result is presented as #NUM! in Excel. In such cases one usually sets $s_{Pos} = 0$ (when variances are pooled the negative values should be used).

In the commutability assessment example, we have 5 RMs and have found significant position effects for RM2 and RM 3 at the 5% significance level for method x, and for RM2 at the 5% significance level and RM1 at the 1% significance level for method y. The question is whether the size of the position effects depends on the RM. The fact that significant position effects have been obtained for some RMs but not for all does not prove that the position effects are different for different RMs (not rejecting a hypothesis does not prove that it is true). As there is no reason to believe that the position effects are different for different RMs we have calculated pooled estimates of the standard deviations (s_{Pos}, cells N48 and S48; and $s_{Pos\text{-}mean}$, cells L48 and Q48) from all RMs.

7. Summary of error components (Sheet Error_comp)

Pooled estimates s_{Pos} determined from the RMs are used as estimates of position effects within a run. Note that s_d corrected for position effects is calculated in the table, denoted $s_{d(corr)}$. The corrected value is the SD of sample specific differences without position effects and should be used if we want to estimate the contribution to the uncertainty of the commutability from sample specific differences alone (not including position effects) that is typically the information of interest to assess suitability of the commutability decision for a given MP.

The estimates of the standard deviation within triplicates, s_e, are larger for RMs than for CSs (*F*-test). Possible causes for this difference may relate to minor differences in the molecular forms of the measurand (e,g, conjugated bilirubin in CSs is a di-glucuronide while in RMs is typically a di-tauro) or the matrix that affect the measurement procedures differently. The estimates s_e from the CSs are taken as representative of the performance of the MSs *x* and *y* respectively.

Explanations to the *F*-test in the sheet *Error_comp*:

A *F*-test is used to test the hypothesis that the SD between replicates is the same for CSs and RMs. *F* is calculated as

$$F = \frac{s^2_{e(CS)}}{s^2_{e(RM)}} \text{ if } s_{e(CS)} > s_{e(CS)} \text{ or } F = \frac{s^2_{e(RM)}}{s^2_{e(CS)}} \text{ if } s_{e(RM)} > s_{e(CS)}.$$

With triplicates of 49 CSs the number of degrees of freedom for $s_{e(CS)}$ is 49·(3-1) = 98; with triplicates of 5 RMs in 5 positions the number of degrees of freedom for $s_{e(RM)}$ is 5·5·(3-1) = 50. The critical value of *F* is found in a table of the *F*-distribution with 98 degrees of freedom in the numerator and 50 degrees of freedom in the denominator in the first case and the reverse in the second case. In this example, we have the second case. If the calculated value of *F* is greater than the 97.5-percentile there is a significant difference at the 5 % level. The reason why the significance level is 5 % and not 2.5 % is that we have decided to have the largest observed variance in the numerator when we calculate *F*.

8. Difference plot for CSs and RMs (Sheet *CS&RM_Diff*)

The difference plots between methods *y* and *x* for the CSs have superimposed the differences for the RMs. Because there was no trend in the differences, lines for the average bias for the CS (B_{CS}), black line, and the criteria for commutability (C) of the RM, dashed red lines, are shown on the plot. A C value of 0.12 (about 12 % in concentration) was used for this example. The uncertainty [$U(d_{RM})$ from sheet *Commutability*] of the difference in bias between the CS and the RM is shown for each RM as error bars. The uncertainty consists of two components: the uncertainty of the estimate of bias for the CSs and the uncertainty of the estimate of bias for each RM. When the uncertainty interval is inside ($B_{CS} \pm$ C) the RM is commutable, when it is outside ($B_{CS} \pm$ C) the RM is non-commutable and when it overlaps the C limits the result is inconclusive.

9. Commutability Assessment (Sheet *Commutability*)

A commutability criterion value (C) of 0.12 was used for this example (the value C = 0.12 was chosen arbitrarily to illustrate the calculations). Refer to the primary paper and to part 1 in this series for information how to determine a criterion. The table shows calculation of the expanded uncertainty [$U(d_{RM})$] of the difference in bias between the CSs and the RMs. The uncertainty consists of two components: the uncertainty of the estimate of bias for the CSs and the uncertainty of the estimate of bias for each RM. A coverage factor of 1.9 was used to give coverage of about 90 %. Coverage of 90 % was used as it corresponds to a one-sided test at the 5% significance level. In this example, there was no trend in the difference plot and B_{CS}, s_B and $u(B_{CS})$ were used in the columns H, I and J and were the same for each RM. A moving average can be used when there is a trend over a concentration interval.

The plot shows the difference in bias (d_{RM}) between the CSs and the RM and its uncertainty compared to the C limits. Each RM was classified as commutable, non-commutable or inconclusive based on whether the (d_{RM}) $\pm U(d_{RM})$ was contained within the C limits, not within the C limits or overlapped a C limit. The same information is also obtained from the difference plot in Sheet *CS&RM_Diff* except the biases are shown on the x-axis rather than the d_{RM} that is centered at zero.

As the ratio in the last column Q is less than 1, the dominating contribution to the uncertainty of d_{RM} comes from the measurements of the RM. The position effects are of the same size as the sample specific differences, see sheet *Error_comp*, and to reduce the uncertainty more positions would need to be used.