IFCC working group recommendations for correction of bias caused by non-commutability of a certified reference material used in the calibration hierarchy of an end-user measurement procedure

Explanation of the examples 1 and 2 spreadsheets that illustrate the concept for determining and using a correction for non-commutability bias of a certified reference material.

This description and the accompanying spreadsheets provide examples to illustrate the concept and approaches for determining a correction for non-commutability bias of an existing certified reference material (CRM) for use in the calibration hierarchy of an end-user measurement procedure (MP). The examples are intended to assist a reader with understanding the main considerations for developing a correction for non-commutability bias with a low enough uncertainty to be useful in a calibration hierarchy. The experimental design including the number of clinical samples (CSs), number of replications, etc. is chosen to ensure the uncertainty of the correction is within the uncertainty budget for the calibration hierarchy. Each provider of an end-user measurement procedure (e.g. an in-vitro diagnostic (IVD) device manufacturer or a laboratory that develops a measuring system for its own use) is responsible to develop a suitable experimental design, detailed procedures and statistical approaches to determine a correction for non-commutability bias of a CRM to use in the calibration hierarchy of their specific end-user IVD device or measuring system.

One-level CRM Example (refer to spreadsheet "Bias Correction Example 1"):

Background:

In a multiple MP commutability experiment, a one-level certified reference material (CRM-A) was found to meet commutability requirements when measured with MP1. In this same commutability experiment, CRM-A was found not to be commutable with clinical samples (CSs) when measured with MP2. The underlying assumption used for providing a one-level CRM is that MPs produce measurement responses for CSs that are linearly proportional to the amount of measurand in those samples over the measuring interval. Consequently, a one-level CRM along with a zero-concentration measurement, such as a reagent blank, is suitable for establishing metrological traceability of the values assigned by the end-user MP over the measuring interval. This assumption that CRM-A is suitable for establishing calibration over the measuring interval is true for each MP whether CRM-A is commutable with that particular MP or not. This assumption can be verified by the manufacturers of the MPs by using procedures described in Clinical and Laboratory Standards Institute (CLSI) EP6.

Experimental design:

The proposed study design incorporates a sample of 50 CSs measured on both MPs to accommodate CS-to-CS variations. A grouping of 10 CS samples plus a CRM-A sample is used to match each group of CS measurements to a CRM-A measurement. This grouping can be considered as a run. There are 5 such runs on each of two days in order to measure all 50 CSs within each day. Within each run, each sample (CRM-A and CSs) is measured in replicates of 2. Within each MP, the difference between each CS and its matched CRM-A sample measurement (per equation 5 below) is taken using the average of the 2

replicates. Because of the matching, and the assumption of equivalent variability of CS and CRM-A measurements, any variance components outside of the run are factored out. These variance components include run-to-run, day-to-day, instrument-to-instrument, reagent lot-to-reagent-lot, and calibration event-to-calibration event. The differences between MP1 and MP2 are then used to compute the CRM-A non-commutability bias correction factor using the 100 results from each CS (50 CSs for each of two days).

The collected data from the example is presented in tab "Raw Data" with the sequence of measurements noted in columns A through E. The two replicate measurements for each CS are seen in columns G, H, J, K, O, P, R, S over both MPs over both days. The means for each day are in columns I, L, Q, T and the overall means for each MP are in in columns M and U.

The assumption of proportionality is tested in tab "CS proportionality". The CS results (columns C and D) and their ratio (column F) are used to present both a ratio plot versus concentration and a scatter plot of MP2 versus MP1 with fitted line originating at (0,0). Both of these plots also display the difference in MP1 versus MP2 measurements in CRM-A that replicates the non-commutability bias seen in the initial commutability experiment.

Other study designs are possible where testing restrictions apply. For example, a point of care (POC) device using a new cartridge for each measurement is itself a run, so matching CRM-A and CS measurements within a run is not feasible. All means of matching such as operator, period of time, cartridge lot, and device should be explored to reduce the effects of such factors.

Value transfer:

The initial commutability experiment established that the comparator MP1 can use the assigned or "true" value of CRM-A in its calibration hierarchy for the end-user calibrator to assign a value to a CS that is traceable to the assigned value of the CRM-A (noted as a "true" value in this example in tab "Raw Data", cell E63). This true value transfer from CRM-A can also be performed without relying on the stability of the calibration hierarchy of MP1 by measuring both CRM-A and CSs simultaneously in the same runs. In doing so, truth can be transferred from the true CRM-A value to the true CS value using equation 1 that avoids any possible drift in calibration of the comparator MP1.

$$obsCS1 * \frac{trueCRMA}{obsCRMA1} = trueCS \qquad [1]$$

where *obsCS1* is the MP1 measured value of a CS, *obsCRM-A1* is the MP1 measured value of CRM-A and *trueCRM-A* is the assigned value of CRM-A that can be used in the calibration hierarchy of MP1 and all other MPs for which CRM-A was found to be commutable with CSs.

The goal of the example 1 experiment is to determine the correction for non-commutability bias of CRM-A required to allow its use in the calibration hierarchy of MP2. Again, the value of CRM-A and the CSs are measured simultaneously but now on MP2. The desired bias correction is determined by using the true CS value assigned using measurements from MP1. This bias correction is performed by transferring truth from the true CS value to the bias corrected CRM-A2 value according to equation 2.

$$obsCRMA2 * \frac{trueCS}{obsCS2} = corCRMA2 \quad [2]$$

where *obsCRM-A2* is the MP2 measured value of CRM-A, *obsCS2* is the MP2 measured value of the same CS whose value (*trueCS*) was assigned using MP1 measurements and *corCRM-A2* is the CRM-A value corrected for non-commutability bias that could be used in the calibration hierarchy for MP2.

Substituting and simplifying provides the correction factor for moving from observed CRM-A2 to a non-commutability bias corrected CRM-A2 value as shown in equation 3.

$$obsCRMA2 * \left(\frac{trueCRMA * obsCS1}{obsCRMA1 * obsCS2}\right) = corCRMA2 \quad [3]$$

Reorganizing these terms puts the equation into a format that groups the observed results together in computing the correction factor for non-commutability bias when CRM-A is used with MP2 as shown in equation 4.

$$\left(\frac{obsCS1}{obsCRMA1}\right) * \left(\frac{obsCRMA2}{obsCS2}\right) = \frac{corCRMA2}{trueCRMA} \quad [4]$$

This equation can be used regardless of whether CRM-A has been used in the calibration hierarchy of either MP or not because CRM-A is measured in the same runs as the CSs using both MPs. The correction factor is derived from the measured values of CRM-A and CSs from each MP. Consequently, the current calibration hierarchy of each MP is not relevant. The only requirement is that CRM-A is commutable for use with the comparator MP1. This equation can be transformed to natural log format with the MP2 measurements expressed in the same order as those from MP1 as shown in equation 5.

The advantage of using natural log transformed data is to eliminate the disproportionate magnitude of a correction factor depending on whether the correction is in a positive or negative direction. For example, for values of MP1 306 and MP2 368, the difference for MP2 vs. MP1 is +20.3% bias. For values of MP1 368 and MP2 306, the difference for MP2 bs. MP1 is -16.8% bias. Using natural log transformed values eliminates this proportional difference in bias.

{ln(obsCS1) - ln(obsCRM-A1)} – {ln(obsCS2) - ln(obsCRM-A2)} = ln(corCRM-A2) – ln(trueCRM-A) [5]

In the equations so far, only a single CS has been measured simultaneously with CRM-A on each MP to make the calculations. Such an experiment could be highly influenced by a CS with a sample specific influence that behaved differently on MP1 versus MP2. Therefore, any CRM-A correction factor experiment should be measured from a reasonable number (e.g. 50 in this example) of CSs to minimize sample specific influences and to identify a CS with a large sample specific influence as an outlier. The calculation of the derived correction for non-commutability bias should be an average over all CSs. This averaging is possible because of the underlying assumption of proportionality of results for CSs over the measuring interval when the MPs have metrological traceability of their end-user calibrators to a single level CRM-A. The experimental design for a one-level CRM selects CSs with values clustered near the value of the CRM to minimize any possible non-proportionality over the measuring interval. As noted above, the proportionality assumption is confirmed in the tab “CS Proportionality”.

Results:

In tab “Log Data” the natural log of the raw data is used with equation 5. The data collection sequence is again in columns A through E. The individual replicates are seen in columns G, H, K, L, Q, R, U, V. Their means for each day for each MP are seen in columns I, M, S, W. The difference between CS and CRM-A for each day are seen in Columns J and N for MP1 and in columns T and X for MP2. The mean difference between CS and CRM-A for MP1 is in column O and the mean difference for MP2 is in column Y. Using equation 5, the non-commutability bias correction based on each individual CS is computed in column AA as the difference of these two average differences.

The non-commutability bias correction for the CRM-A is determined for each of the 50 individual CSs (column AA, rows 5 through 54). The corrected result for CRM-A is shown in column AB in log units and in column AC in reporting units. The mean percent bias in reporting units is seen in cell AC:58, which is the correction factor. The SD of the distribution of individual CS based estimates of corrected CRM-A in reporting units is determined in cell AC:59. The relative uncertainty estimate (CV%) for the mean percent bias (cell AC:58) is shown in cell AC:60. The estimated standard error (SE) of the correction factor for CRM-A is shown in cell AC:61. The relative uncertainty (CV of the SE) of the non-commutability correction factor in cell AC:58 is shown in cell AC:62. Because of the assumption of proportionality, this same uncertainty value is seen at 100 mmol/L (cell AF:62) and 400 mmol/L (cell AH:62) which are concentration values at which a manufacturer may decide to make end-user calibrators.

Once the non-commutability bias correction factor is determined, it can be used to make a correction for non-commutability bias of CRM-A in the calibration hierarchy of MP2. The tab “Calibration Hierarchy” contains an example calibration hierarchy for one of the end-user calibrator levels. Values for the relative standard uncertainty are taken from the certificate for the CRM or from typical values for the working calibrator and the end-user calibrator value assignment steps. The non-commutability bias correction is made to the working calibrators and therefore this correction’s uncertainty is added after the working calibrator value assignment step. The combined relative expanded uncertainty (k=2) at the steps in the calibration hierarchy for the end-user calibrator and for clinical sample results, including the uncertainty of the correction step, are compared to pre-established goals. In this example, adding the correction for non-commutability of the CRM into the calibration hierarchy maintained the combined relative expanded uncertainty within the pre-established goals. The example experimental design and the performance characteristics of the MPs allowed a very small incremental influence of the uncertainty of the correction step. If the added uncertainty of the non-commutability bias correction causes the combined uncertainty goal to be exceeded, a number of steps can be taken to reduce the uncertainty. Such steps include using a larger number of CS, using more replicates of CRM-A interspersed with the CS, or using more days during the experiment.

A number of assumptions and decisions were made in making these example calculations. First, it was assumed that the matching of CRM-A and CS measurements was successful and that the best way to measure the effectiveness of this matching was to compute and thus measure the equation 5 difference estimate for each day and each CS. Second, those factors, such as replicate and day that are not matched are averaged to reduce their variability contribution. Second, the results of equation 5 are untransformed into reporting units before the overall mean, SD and uncertainty are computed. Consequently, the correction factor and its uncertainty added to the calibration hierarchy are expressed

in reporting units or relative reporting units. Other assumptions and decisions regarding the experimental design, matching of measurement conditions and analysis strategy are possible.

This example shows how the concept of a correction for non-commutability bias can be applied and illustrates the main statistical considerations to be included in an experimental design. The specific details of a suitable experimental design will be determined by the IVD manufacturer or clinical laboratory in the case of a laboratory developed test.

Three-level CRM Example (refer to spreadsheet "Bias Correction Example 2"):

Background:

A three level CRM-B is provided for the instance where dilutions of a CRM material do not provide a consistent non-commutability bias compared to CS samples covering the interval to be calibrated. Consequently, a CRM with multiple, individually prepared and certified levels is required for establishing calibration over the measuring interval. For this example, a proportional measurement response is not assumed for CRM-B or CS for either MP5 or MP6.

In a multiple measurement procedure (MP) commutability experiment, a three-level certified reference material (CRM-B_1, CRM-B_2, CRM-B_3) was found to meet commutability requirements for all three levels when measured with MP5. As described in the main report, we assume that MP5 is selected as a comparator MP because its performance in the original commutability assessment was well within the acceptance criterion. In this same commutability assessment, CRM-B was found not to be commutable for any of the levels when measured with MP6.

One way to establish the correction factor for each of the three CRM-B levels for use with MP6 is to run an individual single level experiment for each level as illustrated above in example 1 using CSs that bracket each CRM-B level over a narrow concentration interval.

The example presented below provides an alternate approach to determine correction factors for all three levels of CRM-B from a single experiment. Since proportionality is not assumed then the ratio of CRM to CS measurements also is not assumed to be consistent over the measured interval of the experiment. With CS to CRM ratios potentially changing with different concentrations of CS, the equations 1 through 5 above are no longer applicable and the granularity of matching CRM and CS shifts from within run to within day.

Experimental design:

To characterize the behavior of MP5 and MP6 over a broad measuring interval that includes the concentrations for each level of CRM-B, all samples (both 100 CS and the three CRM levels) are measured interspersed within a sequence over each of two days for both MPs. The number of clinical samples, number of measurements and number of days should be as needed to cover the interval of the CRM levels and to provide adequate uncertainty of the correction to meet the total uncertainty goal for the MP.

The data from the example measurements is presented in tab "Raw Data" with the sequence of measurements noted in column A and the identity of each sample in column B. The two replicate measurements for each sample, CS or CRM-B, are seen in columns D, E, F, G, H, I, K, L over both MPs over both days. The MP6 means for each day are in columns J and M.

In step one, the relationship between the assigned (true) values of a three-level CRM-B and each level's measured values must be characterized. This is done using MP5 for which CRM-B has been shown to be commutable with clinical samples. To do so, measurements of each level of CRM-B are made multiple times over each of two days. In doing so, a mathematical transform is created for each day that can

convert an observed MP5 result to a value that is traceable to the truth as represented by CRM-B. A quadratic fit is used because there are three levels of CRM and, since MP5's calibration hierarchy was created before CRM-B was available, no assumption can be made about the linearity or proportionality that will be observed when CRM-B is measured with MP5. The shape of the quadratic fit between the assigned values of CRM-B_1, CRM-B_2, and CRM-B_3 to their observed values (see equation 6 below) is highly dependent on the calibration of MP5 using its current end-user calibrators, a new calibration event is performed each day before making measurements. Day to day variability therefore encompasses calibration to calibration variability. If such variability is excessive between days, it will be seen by comparing the day 1 and day 2 plots of observed versus true value as shown in the tab "MP5 Analysis". If excessive day-to-day variability is seen then more days, and calibration events, must be included in the experiment. The uncertainty for each CRM-B level at this step is determined by how tight the MP5 CRM-B replicate predictions are around the CRM-B assigned value, using the quadratic fit (true value versus MP5 mean observed value).

In step two, this quadratic relationship (mathematical transform) is used, within each day, to calculate the true values of 100 CS that are again traceable to CRM-B. The CS values are spread across the measuring interval of MP5 such that all three CRM-B levels are overlapped by CS values (see equation 7 below). No CS value should be close to the limit of quantitation of MP5 since no measurements below this value, by definition, provide acceptably accurate results (CLSI EP17). No CS values should be so near the upper end of MP5's measuring interval that any of the results are outside of this interval.

In step three, the true CS values, determined using MP5, are fit with a quadratic equation to the observed values of CS from MP6 for each day (see equation 8 below). In this step, since CS are by definition commutable material, then the quadratic fit provides a relationship (mathematical transform) between MP6 measurements and truth as represented by CRM-B. The uncertainty of this step is determined by the spread of residuals around this new quadratic fit.

In step four, the correction factors for the three CRM-B levels to be used with MP6 are calculated using the quadratic relationship determined in the third step (see equation 9 below). In this case the differences between the predicted values using this relationship and the observed MP6 values provide the corrections required by MP6 if it is to use CRM-B in its calibration hierarchy and get the same CS values of other MPs for which CRM-B is commutable. The uncertainty of this step for each CRM-B level is computed in a similar way as in step one above, by determining how tight the MP6 CRM-B replicate predictions are around the CRM-B corrected value, using the quadratic fit (corrected value versus MP6 mean observed value).

In order to ensure that the CRM-B and the CS measurements are made under similar circumstances, the samples should be interspersed throughout each day. Because of the wide interval of CS concentrations and since a consistent ratio between CS and CRM cannot be assumed across this interval, no effort has been made to recreate the one-level CRM example's run structure to achieve within run CS to CRM matching. However, it is still a reasonable design to use since there is still within-day matching. Alternatively, a random ordering of CRM-B and CS can be used across each day. Random ordering would also reduce any effects of run position or run order.

All of these considerations also apply for measurements using MP6. Again, within each day 100 CS measurements are interspersed with multiple measurements of CRM-B. Since it is likely that a manufacturer will have access to multiple instruments and reagent lots, it would be reasonable to use a different instrument and reagent lot each day for MP6. As for MP5, a new calibration event for each MP6 instrument and reagent lot is required for each day. These three processes (interspersing CS and CRM-B measurements, measurements on two days, calibration of each measuring system each day) ensure that the MP6 end-user calibration process is optimized to provide consistent estimates of the non-commutability bias corrected CRM-B values.

Any residual effects from extra-day, extra-calibration variance components will be included in the directly measured day-to-day imprecision of the correction factor estimates and therefore increase their uncertainty.

Value transfer:

In the three-level CRM-B example, all samples (the 100 CS and the three CRM levels) are measured interspersed within each of two days. Proportionality between the MPs over the measuring interval cannot be assumed and therefore equation (1) cannot be used. Instead, an initial fit of observed CRM-B versus true CRM-B values must be performed through the three levels to cover the measured interval. Given the three log transformed levels, a quadratic equation (equation 6) can be directly solved to provide a true value from the observed values and thus characterize this relationship

$$a_1 * \ln(obsCRMB5)^2 + b_1 * \ln(obsCRMB5) + c_1 = \ln(trueCRMB) \quad [6]$$

where *obsCRM-B5* is the MP5 observed value for one of the CRM-B levels and *trueCRM-B* is the certified value of that CRM-B level that can be used in the calibration hierarchy of MP5 and all other MPs for which CRM-B was found to be commutable with clinical samples. The parameters of this initial quadratic fit can be directly derived since there are three unknowns (a_1, b_1 and c_1) and three equations (one for each CRM-B level).

Using this directly computed relationship for the CRM-B levels in equation 6, true values of all CS samples can now be calculated from their values measured using MP5 using equation 7,

$$a_1 * \ln(obsCS5)^2 + b_1 * \ln(obsCS5) + c_1 = \ln(trueCS) \quad [7]$$

where *obsCS5* is the MP5 observed value for a CS and *trueCS* is the result for that CS whose value is traceable to the certified (true) CRM-B values using the coefficients from equation 6. This set of calculations can be performed each day using the measured result from MP5.

The next step is similar to equation 2 where the calculated true CS values are compared to the MP6 observed values for the same CS specimens.

$$a_2 * \ln(obsCS6)^2 + b_2 * \ln(obsCS6) + c_2 = \ln(trueCS) \quad [8]$$

where *trueCS* is the CS value assigned using results from MP5 and *obsCS6* is the CS value observed using MP6. The three parameters of this second quadratic fit are a_2, b_2 and c_2.

Solving the quadratic equation and substituting CRM-B values for CS values

$$a_2 * \ln(obsCRMB6)^2 + b_2 * \ln(obsCRMB6) + c_2 = \ln(corCRMB6) \quad [9]$$

where *corCRM-B6* is the commutability bias corrected value of each of the three CRM-B values to be used in the calibration hierarchy of MP6 and *obsCRM-B6* is the MP6 observed value of each CRM-B level. Again, each of these corrected CRM-B6 level estimates is calculated for each day.

Results:

In tab "MP5 Analysis" the MP5 data from "Raw Data" are log transformed into individual replicate columns D, E, H, I for both CS and CRM-B. The CRM-B measured values are averaged in cells F:107, F:113, F:119 for Day 1 and cells J:107, J113, J119 for Day 2. The quadratic fit between these observed CRM-B values for each day and the true CRM-B values from the certificate of analysis (in cells B:126, B127, B128) are computed (per equation 6) and the three quadratic parameters are found in cells E:129, E:130, E131 for day 1 and cells I:129, I130, I131 for day 2. Using equation 7, this log-log quadratic relationship is used, within each day, to calculate the true values of 100 CS (see column G for day 1 and column K for day 2). These two are averaged for each CS in column L. Next these quadratic fits, along with the observed MP5 replicates produce predicted replicates for the three CRM-B levels in E:138 through E:167 for day 1 and I:138 through I:167 for day 2. Summary statistics of the predicted replicates are provided in columns F and J for days 1 and 2 respectively. The (observed, predicted) points are plotted for day 1 and day 2 at cells P:137 and P:154. Finally, the uncertainty of the true value predictions at each CRM-B level are determined in rows 172 through 176 for each day and in combination over both days.

In tab "MP6 CS fit", the true CS values determined from MP5 (columns E, K) for each CS are plotted (under column U) against the mean MP6 observed values for each CS for each day (columns D, J). Quadratic equation 8 is fit through this relationship for each day with the parameters shown on each scatterplot and in cells U:6 through U:8 for day 1 and cells U:35 through U:37 for day 2. The predicted value for each CS is determined using these equations for day 1 (column F) and day 2 (column L). The fit residual (columns G and M) and the relative fit (columns H and N) are determined for both days. Finally, the day to day differences are characterized by determining the mean log value of both days (column P), the difference in the days' log quadratic fits (column Q), and the relative ratio of these fits between the days (column R).

The residuals of the observed versus predicted fits are plotted at cells Z:9 and Z:38 for the day 1 and day 2 respectively. The residual of day 1 versus day 2 is plotted at Z:63. All three residual plots indicate that variability doesn't change with change in concentration. Therefore, the uncertainty is calculated as the relative standard error over all results for both within day and between day and combined, as shown in rows 106 through 119. If the residuals are not consistent over the concentration interval, a profile of the uncertainties over the measured interval can be generated and used to estimate the uncertainties at each CRM-B level.

In tab "MP6 Analysis" the quadratic equation parameters from "MP6 CS fit" are replicated in cells D:25 through D:27 for day 1 and H:25 through H:27 for day 2. Observed MP6 CRM-B values (cells C:31

through C:60 for day 1 and cells G:31 through C: 60 for day 2) are used in equation 9 to predict the bias corrected value of each CRM-B replicate in cells D:31 through D:60 for day 1 and cells H:31 through H:60 for day 2. The variability of these predicted values is used to estimate the within day variability of the bias correction at this stage of the value correction process. The (observed, predicted) points are shown, along with the quadratic fit, in plots at cells O:29 and O:46 for day 1 and day 2 respectively.

The between day uncertainty for each CRM-B level is calculated in cells F:64, F:66 and F:68. The bias corrected MP6 CRM-B values plus the relative within day, between day and combined uncertainties are calculated in rows 75 though 77.

In tab "Final Analysis" the uncertainties from the previous 3 tabs are combined for an uncertainty of the entire value correction process for each of the 3 CRM-B levels. Additionally, the assigned CRM-B value is compared with its value as corrected for MP6 non-commutability bias. It should be noted that the amount of correction for the CRM-B levels are 22%, 18% and 23% respectively with uncertainties less than around 1%. Since there is no assumption that the three levels are made of the same material, this range of corrections can be expected. The experiment was designed with enough CS, replicates and days to ensure low uncertainty of these corrections for each of the three CRM-B levels.

Calibration Hierarchy:

The three CRM-B Hierarchy tabs show proposed calibration hierarchy plots for end-user calibrators at values that match the three CRM-B levels as used for MP6. In each one, the uncertainty of the bias correction is added immediately below the uncertainty of that CRM-B level supplied in its certificate. In this way, a manufacturer will use the corrected CRM-B values just as if they had received those values from the CRM manufacturer. The manufacturer's standard procedures for transferring these values to their own working calibrators and subsequently to the end-user calibrators can be followed. The correction can also be applied at a different position in the calibration hierarchy.

This example shows how the concept of a correction for non-commutability bias can be applied for a three-level CRM and illustrates the main statistical considerations to be included in an experimental design. The specific details of a suitable experimental design will be determined by the IVD manufacturer or clinical laboratory in the case of a laboratory developed test.