Statistical Test for Equivalence in Analysis of Commutability Experiments

T. Keller[1] and S. Weber[2]

[1]ACOMED statistik, Leipzig, Germany, [2]University of Applied Sciences Magdeburg-Stendal, Magdeburg, Germany

INTRODUCTION

Commutability is defined as "**equivalence of mathematical relationships among the results of different measurement procedures targeting the same measurand for a reference material (RM) and for representative samples of the type intended to be measured" (clinical samples CS).**

Recently a guideline describing design and analysis of commutability experiments has been published (CLSI C53-P). However, statistical methods proposed in guideline do not refer to inference in terms of equivalence. False positive as well as false negative proof of commutability could follow. We therefore propose application of equivalence testing for analysis of commutability experiments: Equivalence of RM in relationship to CS is shown by analysis of each pair of investigated methods. To show equivalence, the two-dimensional confidence range of RM measurements has to be inside of a range around CS, whereby the limits of this range are prospectively defined by experts.

After analyzing each pair of methods in step (1), the results have to be summarized in a second step (2). Independent on approach used for the first step, difficulties could arise due to different subgroups of methods for which RM is found to be commutable. This general issue remains to be solved.

METHODS - Statistics

In principle, the analysis of commutability experiments is performed within two steps.

(1) First, data of each pair of investigated methods are investigated. As a result, a RM is found to have or not to have same properties as clinical samples measured by a pair of methods.

(2) A contingency table containing results of all pairwise investigations allows to identify groups of methods where a RM is commutable. Difficulties of analysis of this step are demonstrated in results.

The application of proposed procedure refers to step (1) and is performed for each pair of methods:

(1.1) **Prospective definition of limits:** Limits around clinical samples have to be defined by clinical chemistry experts: A RM is considered as commutable if two-dimensional confidence range (CR) of means is within these limits. These limits can be defined by constant coefficient of variation, constant standard deviation or any other imprecision profile related to CS. It should be mentioned, that these limits can be chosen independent on individual methods, with other words: they are equal for all methods.

(1.2) **Mathematical relationship among clinical samples:** Generalized Deming regression [Martin 2000] is used to define relationship among CS measured by two methods. Prediction interval of regression line (results not shown) could be used to assess quality of CS measurements as well as uncertainty of regression.

(1.3) Combine relationship found in (1.2) with limits defined in (1.1)

(1.4) **Determine 2-dimensional confidence range (CR)** of means of replicates **of RM** measured by two methods.

(1.5) **Investigate relationship of CR and predefined limits:**
- If CR is within limits, commutability of RM for pair of methods can be concluded.
- If CR and limits are overlapping, no decision about commutability is possible (too low statistical power).
- If CR is outside of limits, deviation from commutability can be concluded.

RESULTS

The new approach is appropriate for analysis of step (1) in terms of assessing commutability.

Regarding step 2, the situation might occur, that several clusters of methods are identified for which RM are commutable (but not between the clusters). Such situations are demonstrated and are opened for discussion.

RESULTS Step (1): Proof of commutability of RM for pairs of methods

Fig. 1: Example **Method E vs. Method F**.
Predefined Limits: CV=5% (dashed lines)
Assumptions for RM: CV=2%, N = 6
Legend: **coloured symbols: RM with 95%-CR.**
black: CS. All RM with exception of RM D (red symbol) are shown to be commutable for ME/MF.

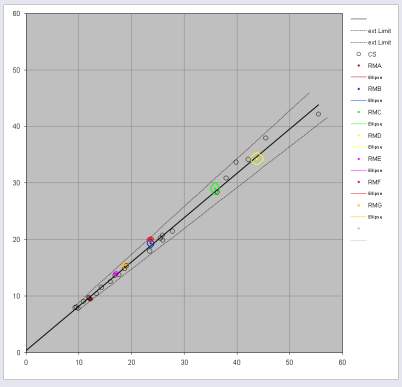
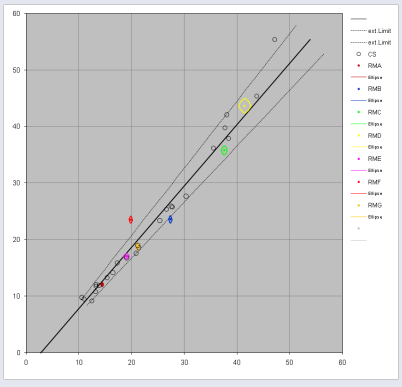


Fig. 2: Example **Method B vs. Method E**.
Predefined Limits: CV=5% (dashed lines)
Assumptions for RM : CV=2%, N = 6
Legend: **coloured symbols: RM with 95%-CR.**
black: CS. All RM with exception of RM A (blue) and RM D (red symbol) are shown to be commutable for MB/ME.



RESULTS Step (2): Summarizing results of step (1) for each RM

RM F: Equivalence test: RM CR vs. Fixed Limits

-	MA	MC	ME	MG	MF	MD	MH	MJ	MB
MA	-	1	1	1	-1	0	0	0	0
MC	1	-	1	1	-1	0	0	0	0
ME	1	1	-	1	-1	0	0	0	0
MG	1	1	1	-	-1	-1	0	0	0
MF	-1	-1	-1	-1	-	0	0	0	0
MD	0	0	0	-1	0	-	-1	-1	0
MH	0	0	0	0	0	-1	-	-1	0
MJ	0	0	0	0	0	-1	-1	-	0
MB	0	0	0	0	0	0	0	0	-

Conclusion: RM F is found to be commutable for one subgroup of methods : A,C,E,G

RM D: Equivalence test: RM CR vs. Fixed Limits

-	MB	MG	MH	MF	MJ	ME	MD	MA	MC
MB	-	1	1	1	1	1	-1	-1	1
MG	1	-	1	1	1	1	-1	-1	-1
MH	1	1	-	1	1	1	-1	-1	-1
MF	1	1	1	-	1	1	1	-1	-1
MJ	1	1	1	1	-	1	1	-1	-1
ME	1	1	1	1	1	-	1	1	-1
MD	-1	-1	-1	1	1	1	-	1	-1
MA	-1	-1	-1	-1	-1	1	1	-	-1
MC	1	-1	-1	-1	-1	-1	-1	-1	-

Conclusion: RM D is found to be commutable within three subgroups of methods : B,E,F,G,H,J; A,D,E and D,E,F,J.

Tables: Examples for contingency table containing results of pairwise comparisons for RM F (left) and RM D (right).
Legend: 1/green: commutable, 0/red: not commutable, -1/yellow: conclusion not possible

METHODS - Data

Data given in CLSI C53-P (7 RM, 25 CS, 9 methods MA-MJ) are used to demonstrate the new statistical method. CS data are very suitable to illustrate different situations which could occur in practice. Unfortunately, CLSI C53-P does contain only means of RM measurements. We therefore simulate replicates of RM measurements with different precision as well as different samples size (here: N=6).

METHODS - Program

A software was developed using MS Excel as environment and software R (www.r-project.org) for statistical analysis.

Within the environment, input of data and parameters is organized, and results delivered from R-program are presented. Software R is used to perform most of known statistical methods [Vesper 2007] for analysis of commutability experiments referring to step (1) (Generalized Deming regression [Martin 2000], Passing Bablok regression, Bland Altman Plots, Linear representation, Correspondence analysis). In addition, the software allows to perform statistical method presented here.

METHODS - two-dimensional confidence range (CR) of means of RM replicates

When RM are measured in independent replicates with two methods, the calculation of two-dimensional CR of the mean is quite complex and has to be performed within numerical computations. The CR follows a contour line of two-dimensional t-distribution assuming no correlation between dimensions. It is diamond shaped for small number of replicates N and is more and more ellipsoid shaped with increasing N. We tabled results of 10000 simulations for meaningful combinations of numbers of RM replicates.

DISCUSSION

We applied equivalence testing on **step (1)** of analysis of commutability experiments. Statistical inference in terms of equivalence is commonly used in guidelines of pharmaceutical industry if similarity of two or more processes within predefined limits has to be proven (e.g. (FDA 2001)). In clinical chemistry literature equivalence testing is mentioned (Lung, Gorko et al. 2003) only seldom, but it is time to introduce these methods. The approach proposed here is advantageous because:

(i) The new method is in accordance with definition of commutability.

(ii) The statistical procedure is simplified.

(iii) Limits of commutability are defined by experts and not by statisticians. In addition, the quality of CS measurements does not influence decision about commutability. Especially usage of prediction interval of Deming regression with variable α-level (!) as proposed in guideline C53-P leads to a dependency of limits on quality of CS measurements on the one hand and an influence of statistician's decision making on the other hand. – Moreover, same limits are used for all comparisons with the new approach.

(iv) Introduction of equivalence testing is connected with a fundamental change of proof of hazard to proof of safety: In conclusion, false positive conclusion of commutability in case of low quality of CS measurements is avoided. On the other hand, false negative non concluding commutability in case of a extremely precise and undistorted measured CS is not possible.

Regarding **step (2)**, situations might occur where several clusters of methods with commutable RM are identified. This issue remains to be solved.

REFERENCES

CLSI EP53-P: Characterization and Qualification of Reference Materials for Laboratory medicine (Proposed Guideline). Clinical and laboratory Standards Institute, Wayne, USA, 2008

Lung, K. R., M. A. Gorko, et al. (2003). "Statistical method for the determination of equivalence of automated test procedures." J Autom Methods Manag Chem 25(6): 123-7.

Martin, R. F. (2000). "General deming regression for estimating systematic bias and its confidence interval in method-comparison studies." Clin Chem 46(1): 100-4.

Vesper, H. W., W. G. Miller, et al. (2007). "Reference Materials and Commutability." Clin Biochem Rev 28(4): 9.

ACKNOWLEDGMENTS

The software was developed for EU Institute for Reference Materials and Measurements (IRMM), Gent, Belgium. We thank Ingrid Zegers and Heinz Schimmel for helpful discussions.