Statistical Test for Equivalence and Non Inferiority in Analysis of Method Validation and Comparison Experiments. Application in Assessment of Carry-Over

T. Keller[1] and T. Brinkmann[2]

[1]ACOMED statistik, Leipzig, Germany, [2]Beckman Coulter Eurocentre, Nyon, Switzerland

INTRODUCTION

A part of method validation (MV) experiments assesses safety to measure same values under special and under normal conditions. In statistical analysis of these experiments, usually the traditional null hypothesis of zero-difference in concentrations is tested. Then, failing to reject the null hypothesis (p>0.05) leads to the conclusion of evidence in favor of safety. **However, absence of evidence is not evidence of absence** (Altman D & Bland, BMJ, 1995; Fig. 1)**.** Therefore conventional approaches like Student's t-test are not appropriate.

Statistical testing of equivalence (two-sided problem) or non inferiority (one-sided problem) is the method of choice. In this report, it is demonstrated for an analysis of a carry-over experiment.

MATERIAL AND METHODS

Carry-over experiment

Serum sample with low ßhCG concentration (L0) was measured 10 times followed by 10 loops of 2 samples of high and 5 samples of low ßhCG concentration samples (L1-L5). The experiments were performed on UniCel® DxC 880i (Beckman Coulter).

STATISTICS

Equivalence of $L1_i$-concentrations (i=1 …10) with other L-concentrations is tested, whereby $L3_i$-$L5_i$ samples are considered. Because carry-over is expected to be associated with an elevated concentration, an one sided test is used (test for non-inferiority).

Abbreviations: CL, (upper) confidence limit; SD, standard deviation; Indices 1 and 3-5 refer to L1 and L3-L5 samples, respectively.

Would you drink water from this glass if the washing process is validated to have no carry-over with p>0.2?

Fig. 1: Absence of evidence is not evidence of absence.

Fig. 1: Carry-over experiment. L, low concentration; H, high concentration
k=10, m=2, n=5, i=10

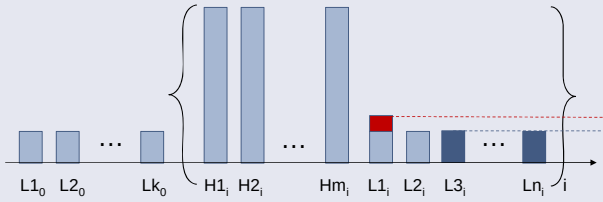
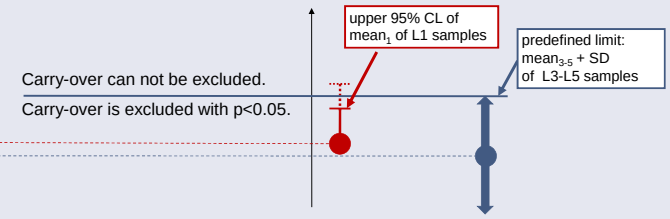


Fig. 2: Principle of statistical test (one sided equivalence test)



RESULTS

For each of the 4 DxC 880i system pipettors P1-P4, null hypothesis was rejected. Equivalence of L1 measurements with L3-L5 measurements and therefore absence of carry-over effect can be concluded.

	P1		P2		P3		P4	
N (L1)	10		10		10		9	
mean$_1$	0.108		1.19		1.18		1.23	
95% CL of mean$_1$	**0.155**	**p<0.05**	**1.25**	**p<0.05**	**1.25**	**p<0.05**	**1.29**	**p<0.05**
N (L3-L5)	30		30		30		27	
mean$_{3-5}$	0.110		1.22		1.20		1.24	
mean$_{3-5}$+SD	**0.191**		**1.31**		**1.32**		**1.32**	

Table 3: Results of analysis. p-values refer to significance of equivalence

Steps	Application: analysis of carry-over experiment
1. Define estimate	Mean$_1$ of $L1_i$ samples.
2. Define limits for equivalence	**Mean$_{3-5}$ + SD** of $L3_i$-$L5_i$ samples.
3. Perform test for equivalence	Check, whether upper 95% confidence limit of estimate as defined in step 1 **is lower** than limit as defined in step 2.

Table 2 How to perform statistical test for equivalence.

Expressed in statistical language, this is: Null hypothesis: **Upper 95% CL of mean$_1$** of $L1_i$-concentrations (i=1...10) is outside of a range described by **mean$_{3-5}$ +SD.** Null hypothesis is rejected on alpha-level=0.05, if upper (one sided) 95%-CL of mean$_1$ is lower than mean$_{3-5}$ + SD. In that case, non inferiority of L1 samples is shown to be significant with p<0.05. (Analysis can be performed with statistical software (here: SAS 9.2) as well as with usual spreadsheet programs, e.g. MS Excel®).

DISCUSSION

Method validation experiments as well as method comparisons assess situations, where values should be proved to be stable or to do not vary within predefined limits. Moreover, it is often necessary to demonstrate equivalence between previously established and new methods. Similar experiments or studies in pharmaceutical or preclinical research (e.g. bioequivalence) are analyzed via statistical methods like tests for equivalence or non inferiority, and those tests are requested by guidelines. Although equivalence tests were introduced in clinical chemistry and laboratory medicine, those statistical methods are not used in daily routine. In clinical chemistry literature, traditional statistical testing (e.g. Student's t-test) is used. A missing significance (p>0.05) is then used as a proof for equivalence, robustness, missing carry-over etc. However, these approaches are not correct, because a too low power (too low sample size) could be the reason for such results.

Statistical equivalence testing should be the statistical method of choice in analyses of experiments, when values should be validated to be stable or to do not vary, e.g. carry-over, robustness, stability, comparisons etc. A panel of statistical methods exists. Relationship of the confidence interval of mean of samples to predefined limits (mean + SD of c of normal samples) is a simple and suitable approach for analysis of common method validation experiments. **By using statistical equivalence testing it was possible to determine that Beckman Coulter's new consolidated instrument UniCel® DxC 880i shows no carry-over.** Finally, changing from non appropriate "conventional" statistical testing to equivalence test is connected with a change of paradigm from proof of hazard to proof of safety.

EuroMedLab 2009, Innsbruck, Austria, June 7-11, 2009