Journal of Chromatography B, 877 (2009) 2208–2213

Contents lists available at ScienceDirect

Journal of Chromatography B

journal homepage: www.elsevier.com/locate/chromb

Review

Review of statistical methodologies used to compare (bio)assays☆

Walthère Dewé

GSK Biologicals, Rue de l'Institut 89, B-1330 Rixensart, Belgium

ARTICLE INFO

Article history:
Received 1 October 2008
Accepted 21 January 2009
Available online 29 January 2009

Keywords:
Assay
Equivalence
Acceptance limits
Confidence interval
Tolerance interval
Probability

ABSTRACT

At any phase of the development of a drug or a vaccine, scientists need to compare different (bio)assays. The objective of this paper is to review different statistical methodologies that could be used to assess the equivalence between methods. Depending on the objective (average or individual equivalence), depending on the design, several approaches are detailed: two one-sided *t*-tests, concordance correlation coefficient, limits of agreement, tolerance intervals and probabilistic approach.

Contents

1. Introduction

At any phase of the life of a drug or a vaccine, scientists need to compare different (bio)assays, two most of the time. As examples:

- A new assay, aimed at describing a biological response and expected to have better characteristics than an assay of reference (higher throughput, cheaper, etc.), is compared to the reference assay.
- A device used in clinical chemistry that should be replaced by a new one.
- When an element of a (bio)assay is modified, the assay in the new settings is compared to the assay in the old settings.
- A (bio)assay that should be transferred from a laboratory to another one, etc.

☆ This paper is part of a special issue entitled "Method Validation, Comparison and Transfer", guest edited by Serge Rudaz and Philippe Hubert.

E-mail address: walthere.dewe@scarlet.be.

1570-0232/$ – see front matter © 2009 Elsevier B.V. All rights reserved.
doi:10.1016/j.jchromb.2009.01.027

In each of these examples, the objective is to demonstrate that the new (bio)assay or the (bio)assay in the new settings generates results that are comparable to those obtained by the old (bio)assay or by the (bio)assay in the old settings. We could talk about equivalence.

Equivalence could be demonstrated at different levels, depending on the context and on the objective of the (bio)assay of interest. Indeed, equivalence should be demonstrated on average for some (bio)assays, while for others, equivalence should be demonstrated on the individual results.

The objective of this paper is to review different statistical methodologies that could be used to assess the equivalence between methods.

2. Study design, assumptions and warnings

2.1. Design

Let us consider a study aiming at assessing equivalence of two (bio)assays.

It is recommended to include in the study design as many factors or sources of variation as possible [1–3]. The rationale is to avoid discovering any issue in the routine application of the new (bio)assay. Theses factors could be operator, equipment, time, result level, etc.

The results generated by both (bio)assays could come from the same set of samples analyzed by both (bio)assays or from two sets of samples having the same origin and identically prepared, a set being analyzed by a (bio)assay and the other being analyzed by the second (bio)assay. The former corresponds to a paired case, the latter to an unpaired case.

The objective of the paper is not to review all the designs that could be envisaged in an assay comparison study. So we will focus on the easiest case where no other factor than the assay is considered. Should multiple factors be included in the study, it is recommended to collaborate with a statistician to elaborate the most appropriate study design and model to maximize the chances to reach the objective while taking into account some possible constraints (e.g. material availability, time, costs, etc.).

2.2. Notation

The number of results generated by assay i ($i = 1, 2$) is noted N_i; the results generated by both assays are noted $\{X_{11}, X_{12}, \ldots, X_{1N_1}\}$ and $\{X_{21}, X_{22}, \ldots, X_{2N_2}\}$. In case of paired case, the number of results generated by both assays is the same ($N_1 = N_2 = N$) and we assume that X_{1i} is paired to X_{2i}, $i = 1, \ldots, N$.

2.3. Normality assumption

We assume that the results generated by each assay are normally distributed:

$$X_{1i} \sim N(\mu_1, \sigma_1^2)$$
$$X_{2i} \sim N(\mu_2, \sigma_2^2)$$

where μ and σ^2 are the parameters of the normal distribution representing the mean and the variance (dispersion around the mean), respectively.

In a paired case, the results generated by both assays are correlated and we note ρ the correlation coefficient. If the results are not paired, we could assume the absence of correlation between them ($\rho = 0$).

In case the normality assumption could not be made, an option is to transform the data to normalize them and then to apply the foreseen methodology on the transformed results. For continuous variables, a frequently used function is the (natural) logarithm when the distribution of the data is positively skewed.

Note that it is possible to identify the most appropriate transformation to be used [4,5].

2.4. Test for difference: inappropriate approach

Any methodology that would be based on a test including the equivalence in the null hypothesis is not appropriate in the context of assay comparison [1–3,6–10]. As a consequence, it is recommended not to use it or to cautiously consider its result as indication only.

Let us consider an example to explain why it is not relevant. Let us assume that the objective is to demonstrate that two assays generate comparable results on average and that the two-sample t-test is performed, i.e. considering on one hand the equality of the means as null hypothesis (H_0) and on the other hand their difference as alternative hypothesis (H_A):

$$H_0 : \mu_1 = \mu_2$$
$$H_A : \mu_1 \neq \mu_2$$

Non-equivalence is concluded if the p-value is lower than the significance level α, or equivalently if the $100(1-\alpha)\%$ confidence interval of the difference between the means does not contain 0.

The length of the confidence interval depends on the number of results. On one hand, in case of a non-(analytically or biologically) relevant difference and large sample size, the confidence interval could not include 0 and we could conclude to non-equivalence. On the other hand, in case of a (analytically or biologically) relevant difference and small sample size, the confidence interval could include 0 and we could conclude to equivalence.

Similarly, the approach that consists, after regressing X_{2i} over X_{1i}, in comparing the intercept and the slope to 0 and 1 [11,12], respectively, is aimed at assessing the difference and is not recommended to evaluate the equivalence.

2.5. Correlation: inappropriate approach

Considering the Pearson correlation coefficient as the lonely statistics to assess an assay comparison is not appropriate either [2,13]. Indeed, the results obtained by two assays could be highly correlated with a systematic difference between them. Moreover, the range of the results also leverages the value of the correlation coefficient: the higher the range, the higher the value of the correlation coefficient.

3. Equivalence on average

Let us suppose that the objective is to demonstrate equivalence on average only.

The first elements to specify, before generating the data, are acceptance limits λ_1 and λ_2, describing the highest difference between the (bio)assays that can be considered as not relevant on an analytical or biological point of view. Frequently, the limits are symmetric: $\lambda_1 = -\lambda_2$.

The approach consists in performing two one-sided t-tests [1–3,5–9]:

$$\text{Test 1}: \begin{cases} H_{01}: & \mu_1 - \mu_2 \leq \lambda_1 \\ H_{A1}: & \mu_1 - \mu_2 > \lambda_1 \end{cases}$$

$$\text{and} \quad \text{Test 2}: \begin{cases} H_{02}: & \mu_1 - \mu_2 \geq \lambda_2 \\ H_{A2}: & \mu_1 - \mu_2 < \lambda_2 \end{cases}$$

Considering a type I error rate α, equivalence is concluded if the $(1-2\alpha)\%$ confidence interval of the difference between the means

is within the interval $[\lambda_1,\lambda_2]$, the confidence interval being given by:

$$(\hat{\mu}_1 - \hat{\mu}_2) \pm t_{\mathrm{df};1-\alpha}\hat{\sigma}_{\mu_1-\mu_2}$$

when $\hat{\mu}_1$ and $\hat{\mu}_2$ are the estimates of the mean results of both assays, $\hat{\sigma}_{\mu_1-\mu_2}$ is the standard error of the difference between the means and df the degrees of freedom associated to the standard error. The details to calculate parameter estimates, standard error and degrees of freedom can be found in Appendix A ([(E.1–E.4, E.6, E.7)]).

Despite the fact that a $(1-2\alpha)$% confidence interval is used to take the decision, the type I error rate of this test is α. Indeed, both null hypothesis could not be true: if the difference is smaller than λ_1, it could not be larger than λ_2, and vice versa. As a consequence, performing the tests 1 and 2 at the α level consists in comparing $(\hat{\mu}_1 - \hat{\mu}_2) - t_{\mathrm{df};1-\alpha}\hat{\sigma}_{\mu_1-\mu_2}$ and $(\hat{\mu}_1 - \hat{\mu}_2) + t_{\mathrm{df};1-\alpha}\hat{\sigma}_{\mu_1-\mu_2}$ to λ_1 and λ_2, respectively, i.e. comparing the $(1-2\alpha)$% confidence interval to $[\lambda_1,\lambda_2]$.

4. Equivalence on individual results

If the objective is to demonstrate the equivalence on individual results, the recommendation is to design the study to obtain paired results. On one hand, it is consistent with the objective and on the other hand, it makes the analysis easier. So let us assume, in this section, that the results are paired.

Different methodologies are available to assess equivalence, depending on the study design:

- Concordance correlation coefficient [13]. To be used only if the samples are independently selected from the same population. In other words, it cannot be used if different clusters exist in the samples that are selected or prepared for the study.
- Bland–Altman [2,14–17]. To be used if the samples are not independently selected from the same population. For example, if samples are prepared at different levels of concentration, the Bland–Altman approach could be used while the concordance correlation coefficient could not be calculated.
- Probabilistic approach [3,8]. To be used in the same context as Bland–Altman.

4.1. Concordance correlation coefficient

The concordance correlation coefficient indicates how the data are distributed around the 45° line through the origin, called the concordance line, if we plot the results obtained with assay 1 against those obtained with assay 2.

The concordance correlation coefficient is estimated as follows (the formula to estimate the Pearson correlation coefficient is given by [(E.5)] in Appendix A):

$$\hat{\rho}_c = \frac{2\hat{\rho}}{(\hat{\sigma}_1/\hat{\sigma}_2) + (\hat{\sigma}_2/\hat{\sigma}_1) + ((\hat{\mu}_1 - \hat{\mu}_2)^2/\hat{\sigma}_1\hat{\sigma}_2)}$$

The concordance correlation coefficient $\hat{\rho}_c$ varies between −1 and 1. In case of a perfect equivalence, i.e. all the points on the concordance line, it is easy to check that $\hat{\rho}_c$ is equal to one. In case of non-equivalence, $\hat{\rho}_c$ is much smaller than 1 or even close to 0 in case of absence of correlation ($\hat{\rho} = 0$).

Equivalence is concluded if the observed concordance coefficient is higher than an equivalence limit c, fixed before generating the data.

To illustrate the use and the meaning of the concordance correlation coefficient, four different examples of device comparison are represented from Figs. 1–4. The calculated concordance correlation coefficients are 0.95, 0.75, 0.17 and 0.07, respectively (see Table 1). Considering $c = 0.7$ as threshold, equivalence is concluded for the examples illustrated in Figs. 1 and 2. Despite the systematic difference that we can observe in Fig. 2, it is small enough to consider that the results generated by both assays are comparable. In the third case, equivalence cannot be concluded due to an excess of variability with the second device. In the last case, the explanation of the non-equivalence is the lack of correlation.

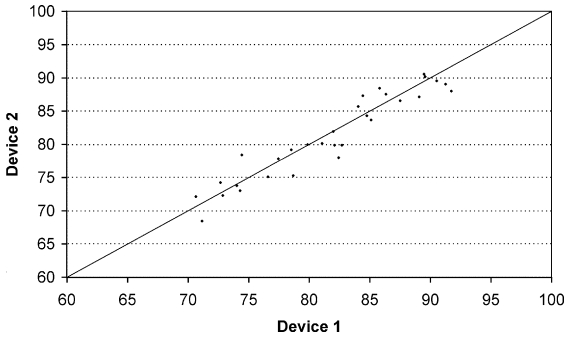

Fig. 1. Device 1 versus device 2 plot with comparable results.

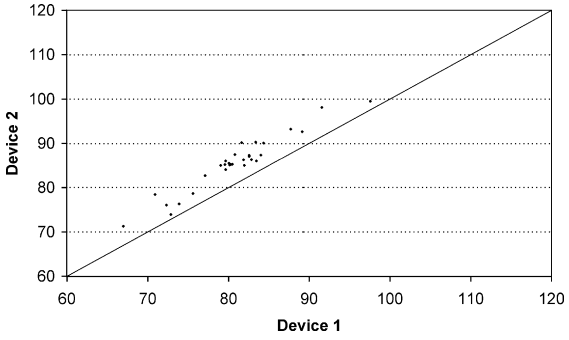

Fig. 2. Device 1 versus device 2 plot with a systematic difference.

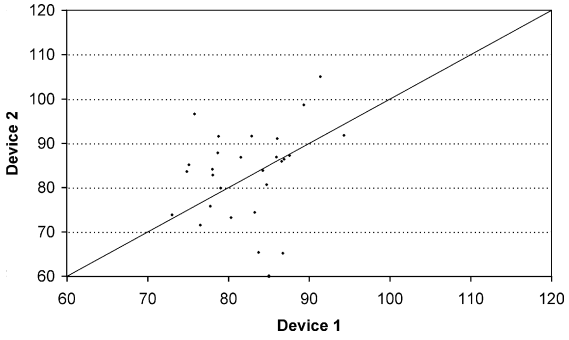

Fig. 3. Device 1 versus device 2 plot with difference in variance.

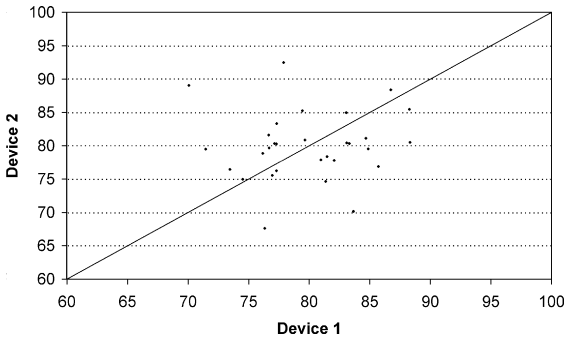

Fig. 4. Device 1 versus device 2 plot with no correlation.

Table 1
Summary statistics of the device comparison examples.

	Example 1		Example 2		Example 3		Example 4	
	Device 1	Device 2	Device 1	Device 2	Device 1	Device 2	Device 1	Device 2
$\hat{\mu} \pm \hat{\sigma}$	81.7 ± 6.39	81.3 ± 6.40	80.8 ± 6.12	85.5 ± 6.33	82.2 ± 5.21	81.0 ± 13.0	79.9 ± 4.74	80.0 ± 5.19
$\hat{\rho}$	0.95		0.96		0.25		0.07	
$\hat{\rho}_c$	0.95		0.75		0.17		0.07	
Two one-sided *t*-test	[−0.18,1.07]		[−5.26,−4.20]		[−2.66,5.23]		[−2.19,2.01]	
Limits of agreement	[−3.52,3.77]		[−8.10,−1.91]		[−23.6,22.1]		[−13.4,11.0]	
Tolerance interval	[−3.77,4.65]		[−8.30,−1.16]		[−25.2,27.7]		[−14.2,14.0]	
Probability	0.001		0.059		0.562		0.278	

4.2. Bland–Altman

The Bland–Altman approach could be considered as visual as well as inferential. Indeed, a so-called Bland–Altman plot could be used as a visual tool to assess the assay comparison, and limits of agreement could be used to decide on the equivalence of both (bio)assays by comparing them to acceptance limits.

The Bland–Altman plot represents, on the *Y*-axis, the individual differences $D_i = X_{1i} - X_{2i}$ ($i = 1, \ldots, N$) versus the mean results $\bar{X}_i = (X_{1i} + X_{2i})/2$ on the *X*-axis. If both assays are comparable, the individual differences should be randomly distributed around the 0 horizontal line, whatever the result level.

In addition to the visual inspection, limits of agreement, giving information about the distribution of the individual differences, could allow us to decide whether two (bio)assays are comparable. Before taking a decision, we have to identify acceptance limits describing the highest tolerable difference between individual results, and the proportion of individual differences that we like to be within these acceptance limits. Similarly to equivalence on average, both assays are considered as comparable if the limits of agreement are within the acceptance limits.

Note that there is no requirement about covering any targeted value (0 for example). Indeed, in concordance with the comments about the inappropriateness of the tests for difference, covering a target is not mandatory as long as the acceptance limits are meaningful. For example, if the limits of agreement do not cover the equivalence target (0 for example) and if the mean difference estimate is within the acceptance limits, equivalence could still be concluded if the variability of the difference is small enough to obtain most of the individual difference within the limits as well.

Assuming the variance of the differences is homogeneous across the range of results, limits of agreement are computed as follows:

$$\hat{\mu}_D \pm z_{1-(\alpha/2)}\hat{\sigma}_D$$

where $\hat{\mu}_D$ is the estimate of the mean individual differences (see formula [(E.8)] in Appendix A for details), $\hat{\sigma}_D$ is the standard deviation of these differences (see formula [(E.9)] in Appendix A for details) and $z_{1-(\alpha/2)}$.the $(1-\alpha)$% percentile of the $N(0,1)$ normal distribution.

In case the assay comparison has to be assessed in a very large range, it is likely that the assumption of variance homogeneity does not hold. In such a case, a common practice is to try a logarithm transformation and compute the limits of agreement on the log transformed data, i.e. the individual difference of the log transformed results. By taking the anti-log of these differences, we obtain the ratio of the actual results. As a consequence, an alternative Bland–Altman plot is used: individual ratio on the *Y*-axis and geometric mean on the *X*-axis.

In order to better control the risk of wrongly concluding equivalence, tolerance intervals could be used instead of the limits in agreement. Indeed, such intervals take into account the (lack of) precision in the estimation of the mean difference as well as the (lack of) precision in the estimation of the variance of the individual differences [18], while the limits of agreement considers only the latter. The former is managed by the inclusion of the standard error of the mean difference in the variance term. The lack of precision in the estimation of the variance is managed by considering a *t*-distribution, with $N-1$ degrees of freedom, instead of the Normal distribution. The tolerance intervals are calculated as follows:

$$\hat{\mu}_D \pm t_{N-1;1-(\alpha/2)}\sqrt{1+\frac{1}{N}}\hat{\sigma}_D$$

where $t_{N-1;1-(\alpha/2)}$.the $(1-\alpha)$% percentile of the Student distribution with $N-1$ degrees of freedom.

Note that these (tolerance) intervals have a limitation in case of a systematic difference between the assays [19,20]. Indeed, in case of bias, the risk to have a bound outside the acceptance limits increases while the proportion of individual differences within the acceptance limits could still be acceptable (the location of the other bound would compensate).

4.2.1. Examples

Let us come back to the four examples already discussed in Section 4.1. Limits of agreement and tolerance intervals are detailed in Table 1, and the Bland–Altman plots are illustrated in Figs. 5–8.

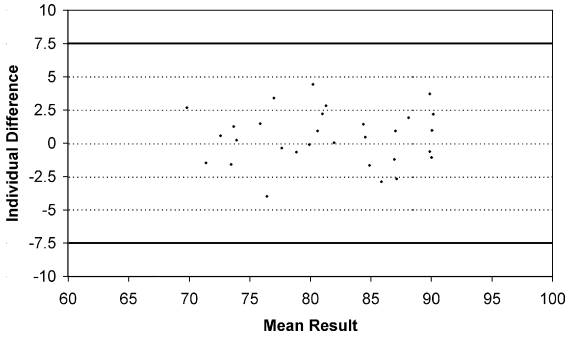

Fig. 5. Bland–Altman plot of the device comparison study with comparable results, where the limits of acceptance are ±7.5 in the result units (solid lines).

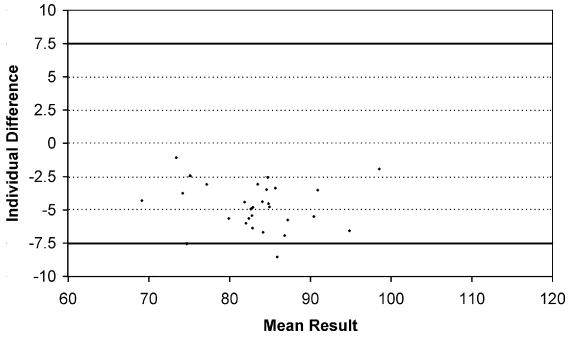

Fig. 6. Bland–Altman plot of the device comparison study with systematic difference, where the limits of acceptance are ±7.5 in the result units (solid lines).

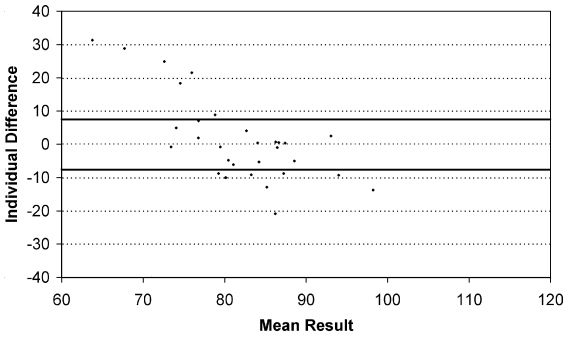

Fig. 7. Bland–Altman plot of the device comparison study with difference in variance, where the limits of acceptance are ±7.5 in the result units (solid lines).

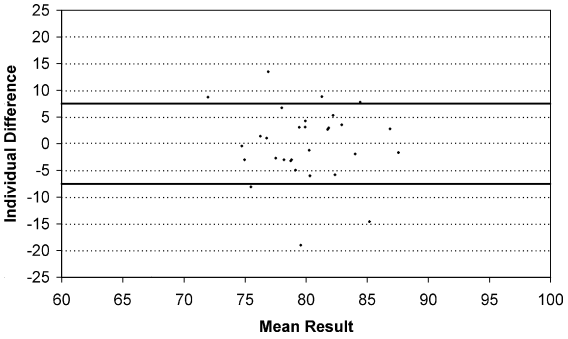

Fig. 8. Bland–Altman plot of the device comparison study with no correlation, where the limits of acceptance are ±7.5 in the result units (solid lines).

Considering ±7.5 as acceptance limits and 0.95 as proportion of individual differences within these limits, only the first example would be successful. Although there is some evidence that both examples 3 and 4 do not represent comparable methods, the example 2 displaying a small systematic difference between the devices is borderline as the lower limit of agreement and the lower bound of the tolerance interval barely missed the target.

Let us consider a fifth example illustrating the comparison of two bio-assays. In this example, the samples are prepared at five different levels of concentration, each being analyzed by each assay. The acceptance limits are ±5 in the result units and 0.95 is targeted as proportion of individual differences within the acceptance limits. The Bland–Altman plot is illustrated in Fig. 9. The limits of agreement are [−4.50,4.05] and the tolerance interval is [−4.66,4.20]. As both intervals are within the acceptance limits, we conclude that both assays have comparable results.

Let us assume that the limits of agreement are within the acceptance limits and that the tolerance interval is not. As the difference between these intervals is only due to sample size, it would be preferable to take a decision based on the tolerance interval. The decision could be to increase the sample size in order to have more information that would possibly give us more evidence about equivalence.

Note that the assumption of homogeneous variance seems reasonable by looking at the Bland–Altman plot.

4.3. Probabilistic approach

In the same philosophy as the methodology detailed in the previous section, the decision to conclude equivalence could be based on a calculated risk of having an individual difference outside our acceptance limits.

The risk corresponds to the probability of having an individual difference out of the acceptance limit, i.e. the sum of the probability that the difference is above the upper acceptance limit and the probability that the difference is below the lower acceptance limit. This risk can be estimated as follows:

$$\begin{aligned}\hat{\pi} &= P(X_1 - X_2 < \lambda_1) + P(X_1 - X_2 > \lambda_2) \\ &= P\left(t_{N-1} < \frac{\lambda_1 - (\hat{\mu}_1 - \hat{\mu}_2)}{\sqrt{\hat{\sigma}_1^2 + \hat{\sigma}_2^2 - 2r\hat{\sigma}_1\hat{\sigma}_2}}\right) \\ &\quad + P\left(t_{N-1} > \frac{\lambda_2 - (\hat{\mu}_1 - \hat{\mu}_2)}{\sqrt{\hat{\sigma}_1^2 + \hat{\sigma}_2^2 - 2r\hat{\sigma}_1\hat{\sigma}_2}}\right)\end{aligned}$$

Both assays are considered as comparable if the probability is small enough, i.e. smaller than a maximum tolerable proportion.

This approach does not have the same border effect than the one linked to the tolerance intervals. Indeed, even if a (single) bound of the tolerance interval barely misses the target due to a systematic

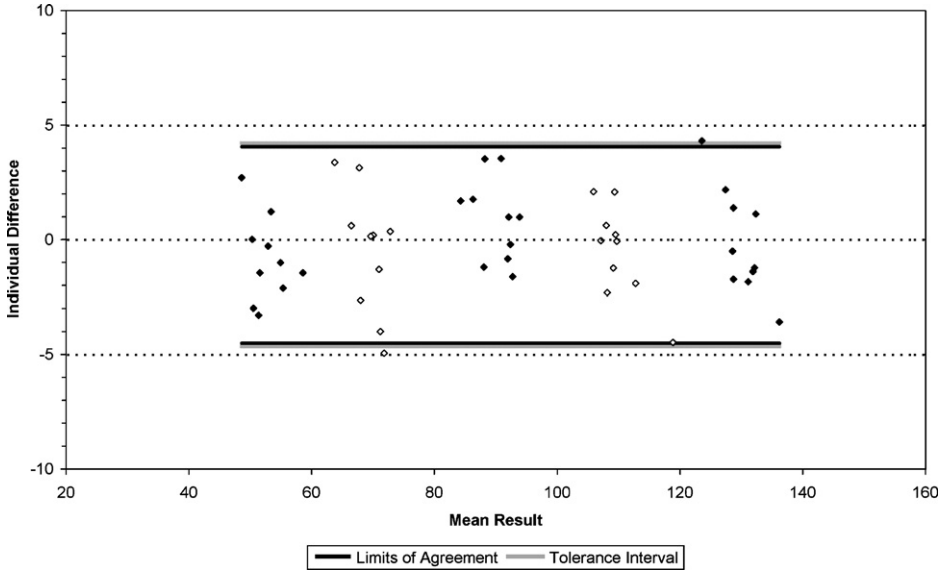

Fig. 9. Bland–Altman plot of an assay comparison study where the limits of acceptance are ±5 in the result units. The successive changes of symbol indicate the different levels of concentration.

error, the probability could be smaller than the tolerable proportion if the opposite tail of the individual difference distribution is largely within the acceptance limits. So while well controlling the risk of falsely concluding equivalence, the probability of success is always higher than the one obtained with a tolerance interval approach when the assays are truly comparable [3].

In case of a multiple levels of concentration, if the assumption of homogeneous variance is valid, the probability $\hat{\pi}$ could be calculated over the whole range. However, if there is evidence that the variance is not homogeneous, the risk can be calculated by level using the variance estimates obtained at each level. In the latter case, the variance could be modeled to gain precision, which requires the help of a statistician.

4.3.1. Examples

Regarding the first four examples, the risk values can be found in Table 1. This probabilistic approach confirms that the observations and conclusions reached with the Bland–Altman analysis.

About the fifth example, the risk to have an individual difference outside the acceptance limits is 0.048. It is small enough, i.e. less than 0.05, to conclude of equivalence.

5. Discussion

Proving the equivalence of two (bio)assays requires specific methodologies. So the Student two-sample *t*-test or any other analogous hypothesis test, although frequently used, are relevant only in the cases where the objective is to detect differences and as consequence, are not appropriate in our context.

Proving equivalence is more exigent and challenging than proving a difference. The difficulties to prove equivalence are the determination of relevant acceptance limit(s) and the fact that the appropriate methodologies require a larger sample size.

A potential key of success of a (bio)assay comparison is to develop a scientific partnership between the scientist in charge of the (bio)assay and the statistician. Both of them have the complementary skills to design the study, to analyze the data and to take a good decision.

If the objective is to assess the equivalence on average, the recommendation is to use the two one-sided *t*-test.

If the objective is to assess the equivalence on individual results, the recommendation is to use the probabilistic approach as, in contrast to the limits of agreement or the tolerance intervals, it does not suffer from the border effect when a systematic difference exists. Regarding the concordance correlation coefficient, fixing a meaningful acceptance limit could be difficult.

Acknowledgements

The author would like to thank Philippe Hubert and Serge Rudaz for giving the opportunity to write a review about method comparison in this special issue.

Appendix A

$$\text{Mean estimate of assay 1}: \hat{\mu}_1 = \frac{1}{N_1}\sum_{i=1}^{N_1} X_{1i} \quad \text{(E.1)}$$

$$\text{Mean estimate of assay 2}: \hat{\mu}_2 = \frac{1}{N_2}\sum_{i=1}^{N_2} X_{2i} \quad \text{(E.2)}$$

$$\text{Standard deviation of assay 1}: \hat{\sigma}_1 = \sqrt{\frac{\sum_{i=1}^{N_1}(X_{1i}-\hat{\mu}_1)^2}{N_1-1}} \quad \text{(E.3)}$$

$$\text{Standard deviation of assay 2}: \hat{\sigma}_2 = \sqrt{\frac{\sum_{i=1}^{N_2}(X_{2i}-\hat{\mu}_2)^2}{N_2-1}} \quad \text{(E.4)}$$

Pearson correlation coefficient (paired case only):

$$\hat{\rho} = \frac{\sum_{i=1}^{N}(X_{1i}-\hat{\mu}_1)(X_{2i}-\hat{\mu}_2)}{(N-1)\hat{\sigma}_1\hat{\sigma}_2} \quad \text{(E.5)}$$

Standard error of the mean difference:

if paired case :

$$\hat{\sigma}_{\mu_1-\mu_2} = \sqrt{\frac{\hat{\sigma}_1^2+\hat{\sigma}_2^2-2\hat{\rho}\hat{\sigma}_1\hat{\sigma}_2}{N}}, \quad (\mathrm{df}=N-1) \quad \text{(E.6)}$$

if unpaired case :

$$\hat{\sigma}_{\mu_1-\mu_2} = \sqrt{\frac{\hat{\sigma}_1^2}{N_1}+\frac{\hat{\sigma}_2^2}{N_2}}, \quad (\mathrm{df}=N_1+N_2-2) \quad \text{(E.7)}$$

Mean estimate of individual differences:

$$\bar{D} = \frac{1}{N}\sum_{i=1}^{N}(X_{1i}-X_{2i}) \quad \text{(E.8)}$$

Standard deviation of individual differences:

$$\hat{\sigma}_D = \sqrt{\frac{\sum_{i=1}^{N}(D_i-\bar{D})^2}{N-1}} \quad \text{(E.9)}$$

References

[1] F. Minois-Offroy, Y. Appriou, V. Brousset, E. Chapuzet, G. de Fontenay, W. Dewé, E. Dumas, C. Ellie, M. Galiay, N. Lefebvre, P. Mottu, M.P. Quint, F. Schoeffter, STP Pharma Pratiques 12 (2002) 337.
[2] Y. Tsong, J. Zhong, K. Lee, Proceedings of the 1st ISBS Symposium, Shanghai, China, June 30–July 2, 2008.
[3] W. Dewé, B. Govaerts, B. Boulanger, E. Rozet, P. Chiap, P. Hubert, Chemometr. Intell. Lab. Syst. 85 (2007) 262.
[4] G. Box, D. Cox, J. Roy. Stat. Soc. Series B 26 (1964) 211.
[5] R.J. Carroll, D. Ruppert, Transformation and Weighting in Regression, Chapman and Hall, New York, 1988.
[6] U. Schepers, H. Wätzig, J. Pharm. Biomed. Anal. 39 (2005) 310.
[7] R. Kringle, R. Khan-Malek, F. Snikeris, P. Munden, C. Agut, M. Bauer, Drug Inf. J. 35 (2001) 1271.
[8] E. Rozet, W. Dewé, R. Morello, P. Chiap, F. Lecomte, E. Ziemons, K.S. Boos, B. Boulanger, J. Crommen, P. Hubert, J. Chromatogr. A 1189 (2008) 32.
[9] D.J. Schuirmann, J. Pharmaco. Biopharm. 15 (1987) 657.
[10] J. Vial, A. Jardy, P. Anger, A. Brun, J.M. Menet, J. Chromatogr. A 815 (1998) 173.
[11] H. Passing, W. Bablok, J. Clin. Chem. Clin. Biochem. 21 (1983) 709.
[12] H. Passing, W. Bablok, J. Clin. Chem. Clin. Biochem. 22 (1984) 431.
[13] L. Lin, Biometrics 45 (1989) 255.
[14] D.G. Bland, J.M. Altman, The Statistician 32 (1983) 307.
[15] D.G. Bland, J.M. Altman, Lancet 8 (1986) 307.
[16] P.J. Twomey, Ann. Clin. Biochem. 43 (2006) 124.
[17] K. Dewitte, C. Fierens, D. Stöckl, L.M. Thienpont, Clin. Chem. 48 (2002) 799.
[18] R.W. Mee, Technometrics 26 (1984) 251.
[19] B. Boulanger, W. Dewé, A. Gilbert, B. Govaerts, M. Maumy, Chemometr. Intel. Lab. Syst. 86 (2007) 198.
[20] B. Govaerts, W. Dewé, M. Maumy, B. Boulanger, Qual. Reliab. Eng. Int. 24 (2008) 667.