Available online at www.sciencedirect.com
SCIENCE DIRECT®
Journal of Pharmaceutical and Biomedical Analysis 39 (2005) 310–314
JOURNAL OF PHARMACEUTICAL AND BIOMEDICAL ANALYSIS
www.elsevier.com/locate/jpba

Short communication

Application of the equivalence test according to a concept for analytical method transfers from the International Society for Pharmaceutical Engineering (ISPE)

U. Schepers, H. Wätzig *

Institute of Pharmaceutical Chemistry, TU Braunschweig, Beethovenstr. 55, D-38106 Braunschweig, Germany

Received 4 December 2004; received in revised form 28 February 2005; accepted 2 March 2005
Available online 31 May 2005

Abstract

The performance of the equivalence test in the context of analytical method transfers was investigated by means of a simulation study. An ISPE design proposal and typical error contributions for pharmaceutical routine control have been used for the testing of accuracy. Acceptable results (probability of a correct decision) have been obtained here. For total variations above 0.4% R.S.D. the basic design was not sufficient. An overview for the number of additional series needed corresponding to higher variations has been developed based on further simulations. An alternative approach may be the choice of wider acceptance criteria, which was also evaluated.

Keywords: Equivalence test; Analytical method transfer; Quality control; Pharmaceutical; Accuracy; Simulation; Acceptance criteria; ISPE

1. Introduction

1.1. Shortcomings of classical approaches compared to the equivalence test

The success of an analytical method transfer for pharmaceutical routine control is tested by comparing statistical results such as means or standard deviations (here only the means are discussed) of the participating laboratories obtained after analyzing samples of the same substance [1,2]. If the test results suggest the rejection of a method transfer, there are two possibilities: the correct rejection of an inappropriate transfer as well as rejecting a good method transfer. The decision has to be made without the knowledge of the true situation. On the other hand, if the results suggest accepting a method transfer, there are again two underlying possibilities connected to this decision. First, the decision to accept the transfer is correct. This is certainly desirable, but the probability for this case as well as the risk, that an inappropriate transfer is accepted, depends on the statistical test used. The latter risk is the more important one. It should be strictly controlled because its consequence would be release/reject decisions based on an inadequate performance of the method in pharmaceutical routine control.

For the evaluation of the performance of a statistical test both risks have to be considered. The classical approach (*t*-test) is only capable of controlling the risk of wrongly rejecting a transfer with an a priori known probability. Therefore a probability value α is chosen. The risk of accepting an inappropriate transfer is only accessible for each special case but not in general. It is evident, that the equivalence test controls this more important risk of passing a bad transfer with an a priori chosen error probability. This feature is the advantage of the equivalence test, which makes it the more suitable test for method transfer situations [3,4]. In order to gain information of the probabilities of the risks connected to the test situations, simulation studies have been carried out, in which virtual transfer situations with known true bias have been generated.

* Corresponding author. Tel.: +49 531 391 2764; fax: +49 531 391 2799.
E-mail address: h.waetzig@tu-bs.de (H. Wätzig).

0731-7085/$ – see front matter © 2005 Elsevier B.V. All rights reserved.
doi:10.1016/j.jpba.2005.03.015

1.2. ISPE design

The International Society of Pharmaceutical Engineering (ISPE) is a global, not-for-profit membership organization that provides education, training, and technical publications to Life Science professionals, including topics of general interest about quality assurance.

An ISPE design proposal for assays of active pharmaceutical ingredients was used. The guide recommends: "that at least two analysts at each laboratory should independently analyse three sample lots in triplicate; resulting in three distinct executions of the method" [1]. The criterion for the performance of the equivalence test was the probability to accept the analytical method transfer, in the case of an acceptably small bias.

According to this design two approaches are thinkable. First, one could understand the term sample lot as sample. Then just three different samples are drawn and each is analysed in triplicate. The variability between the three samples can be considered as being very small in this case. This procedure thus corresponds to a nine-fold application of the method. Consequently only the variation due to the different laboratories is taken into account.

The second understanding considers the term sample lot as sample from a different batch. Using this interpretation, it can be tested if the method is suitable in different laboratories even if the impurity profile slightly differs. Put differently, the selectivity of a method and its robustness is retested.

If all analysts analyse the same three batches, then the amount of additional uncertainty between the batches does not influence the overall probability to accept the equivalence test. However, if the batch variability found differs between the laboratories this cannot be detected without additional measures if the means are identical, e.g. considering batch means of 97, 100 and 103 in one lab and 99, 100 and 101 in the other one. Note, that for the second approach neither the use of a paired *t*-test nor the simultaneous testing of the three batches is suitable, because then the analyst is not an independent error component anymore, leading to correlated results.

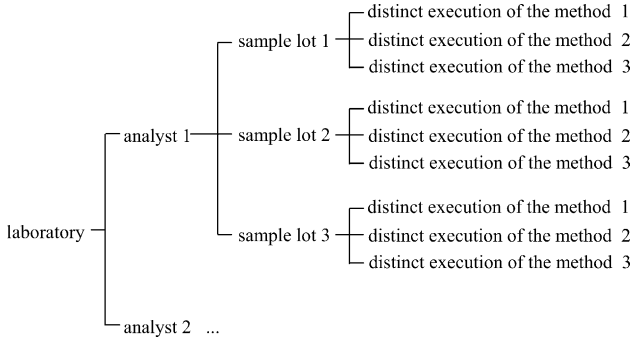

Fig. 1. Basic error structure within a laboratory according to the ISPE design. An analyst independently performs three methods (including sample preparation and measurement) using three sample lots.

Therefore, data sets have been generated according to the first approach. An a priori known error structure was used, considering nine distinct executions of the method for each analyst (Fig. 1).

2. Test procedure

2.1. Assumptions

In a first step the following assumptions have been made: the observed error contributions are independent from each other, both laboratories have the same error structure, the acceptance criteria for the upper and lower limit of the equivalence test were set to typically values of ±2% and the error probability α (one-sided) was chosen to be 0.05 (see Section 2.3). To get a first idea about the performance of the equivalence test under these conditions a typical transfer situation was simulated. For the specific error terms connected to the application of the method the values 0.3% (R.S.D.) for system precision $\hat{\sigma}_{SYS}$ and 0.6% for sample preparation $\hat{\sigma}_{SP}$ have been chosen, which have been evaluated as of a typical magnitude for LC assays [5]. The variation caused by the analysts $\hat{\sigma}_{AN}$ was initially set to 0.3% and varied later (see Section 4). The equivalence test uses the arising mean values of the analysts as single values. Therefore the performance of the equivalence test is observed with respect to the standard deviation of the means of the analysts $\hat{\sigma}_{\bar{x}_{AN}}$ and an a priori known bias. Applying the above mentioned individual error values according to Eq. (1) yields 0.37% for $\hat{\sigma}_{\bar{x}_{AN}}$:

$$\hat{\sigma}_{\bar{x}_{AN}} = \sqrt{\hat{\sigma}_{AN}^2 + \frac{\hat{\sigma}_{SYS}^2 + \hat{\sigma}_{SP}^2}{9}} = \sqrt{0.3^2 + \frac{0.6^2 + 0.3^2}{9}} = 0.37 \quad (1)$$

Note, that different individual error values yielding the same $\hat{\sigma}_{\bar{x}_{AN}}$ have equal probabilities according to the equivalence test.

2.2. Simulation

First, a basic population according to the above mentioned error structure was generated, providing 2000 mean values for each laboratory. Every single result of the basic population was affected by the relative standard deviation of the execution of the method (system precision and sample preparation) and the analysts like described above (see Fig. 1). Then a specified bias of 0 or 1% was integrated between both laboratories (for clearness only two bias-steps are shown). Note that a bias of 2% would be at the limit of acceptance and would provide a probability of 5% to accept the method transfer (corresponding to the chosen $\alpha = 0.05$ of the test, see Fig. 2).

Then the equivalence test was performed (see Section 2.3) using two mean values per site resulting in 2000 tests. The relation of the accepted tests to the number of all performed

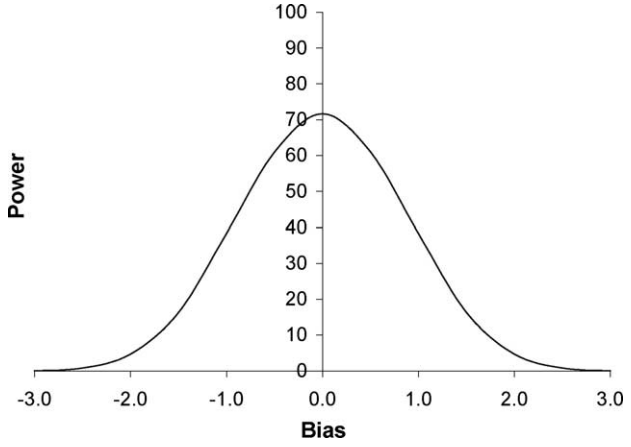

Fig. 2. The principal course of the power against a bias between two laboratories. The power at the limit of acceptance (bias = 2.0) is controlled to 5%.

tests was the probability to correctly accept the transfer. These probability values should be understood with an uncertainty of ±0.5.

2.3. Equivalence test

The percentage 90%-confidence interval of the equivalence test is consisting of a lower limit C_L and an upper limit C_U. Each limit is using a one-sided 95%-confidence interval and it is calculated by Eq. (2) [6]:

$$C_L = 100\left[\left(\frac{\bar{x}_1}{\bar{x}_2}\right) e^{-(t_{\alpha,(2n-2)}\hat{\sigma})} - 1\right],$$

$$C_U = 100\left[\left(\frac{\bar{x}_1}{\bar{x}_2}\right) e^{(t_{\alpha,(2n-2)}\hat{\sigma})} - 1\right] \quad \text{with}$$

$$\hat{\sigma} = \sqrt{\frac{1}{2n}(\hat{\sigma}_1^2 + \hat{\sigma}_2^2)\left(\frac{1}{\bar{x}_1^2} + \frac{1}{\bar{x}_2^2}\right)} \qquad (2)$$

Whereas $\bar{x}_{1/2}$ represent the laboratory means, $\hat{\sigma}_{1/2}$ the corresponding standard deviations. For the calculation of the laboratory means and the corresponding standard deviations in this case the respective mean values of the participating analysts have been used as single values. Further, the value $t_{\alpha,(2n-2)}$ is derived from the *t*-distribution with n degrees of freedom and the one-sided error probability α.

The transfer was accepted, when the lower and the upper limit of the 90%-confidence interval were completely within the interval of ±2%.

2.4. Software

All calculations were performed by Microsoft® Excel 97 (Munich, Germany). The $N(\mu,\sigma^2)$ normal distributed single values x_i were generated by the function: `norminv(random⟨⟩,μ,σ)`. The mean value μ was always 1.0 and the variance σ^2 (here in terms of standard deviation) like described above (e.g. $\sigma = 0.003 = 0.3/100 = 0.3\%$ for system precision).

3. Results using the ISPE concept

The ISPE recommendations are often interpreted, that three different batches should be used in a method transfer investigation. However, as already mentioned in Section 1.2, the paired *t*-test and the simultaneously testing of three batches are not suitable for this approach, because the analyst is not an independent error source and therefore correlation is produced.

The extent of correlation and its effect on the results simultaneously testing three batches was studied. A simulation with $\hat{\sigma}_{AN} = 0.3\%$, $\hat{\sigma}_{SP} = 0.6\%$, $\hat{\sigma}_{SYS} = 0.3\%$ and a bias of 2% lead to a probability of accepting the equivalence test of 9%. Hence the equivalence test is not able to control the risk of passing a bad transfer to 5%. Moreover, the use of a paired *t*-test for this approach was investigated. The values $\hat{\sigma}_{AN} = 0.5\%$, $\hat{\sigma}_{SP} = 0.6\%$, $\hat{\sigma}_{SYS} = 0.3\%$ and a bias of 2% have been implemented in a further simulation. The result was a probability of accepting the test of 19%, which is quite higher as the required 5%, too.

Therefore the initial simulation was performed considering only one batch and two analysts for each laboratory. The considerations of Section 2.1 have been applied resulting in a variation of the analysts of 0.37%. The arrow (a) in Fig. 3 marks this opening result.

A probability of 89% to correctly accept such a transfer was found to be adequate assuming an error amount of 0.3% for the analysts and no bias. In such a situation this design proposal is suitable. The probability to accept a transfer with a bias of 1% (arrow (b)) is always lower due to the smaller distance to the limit of acceptance (see Fig. 2). For the situation, where $\hat{\sigma}_{\bar{x}_{AN}} = 0.62\%$ (applying Eq. (1) with $\hat{\sigma}_{AN} = 0.58\%$) one can see a clear decrease of the probability (50%). This shows, that this design is not appropriate for comparatively higher error contributions. The probability could be heightened if $N = 3$ analysts were involved as a modification of the design proposal (then the probability increases from 50 to 88%, shown by arrow (c)). A value of 0.3% for the variation caused by analysts could be seen as a quite low contribution. Additional error sources between the series apart from analysts (e.g. different days), which might increase the variation, should be taken into account. This was made in the following section.

4. Extension of the ISPE concept

The ISPE concept has been extended by additionally regarding other sources between the series apart from the error caused by analysts. The term $\hat{\sigma}_{AN}$ was replaced by $\hat{\sigma}_B$ to demonstrate, that not only the variation due to different analysts but also due to other additional error sources is probably included. As one could change the numbers of proposed replicates n_{SYS} or sample preparations n_{SP} the standard deviation of the series-means $\hat{\sigma}_{\bar{x}_{series}}$ is given by the

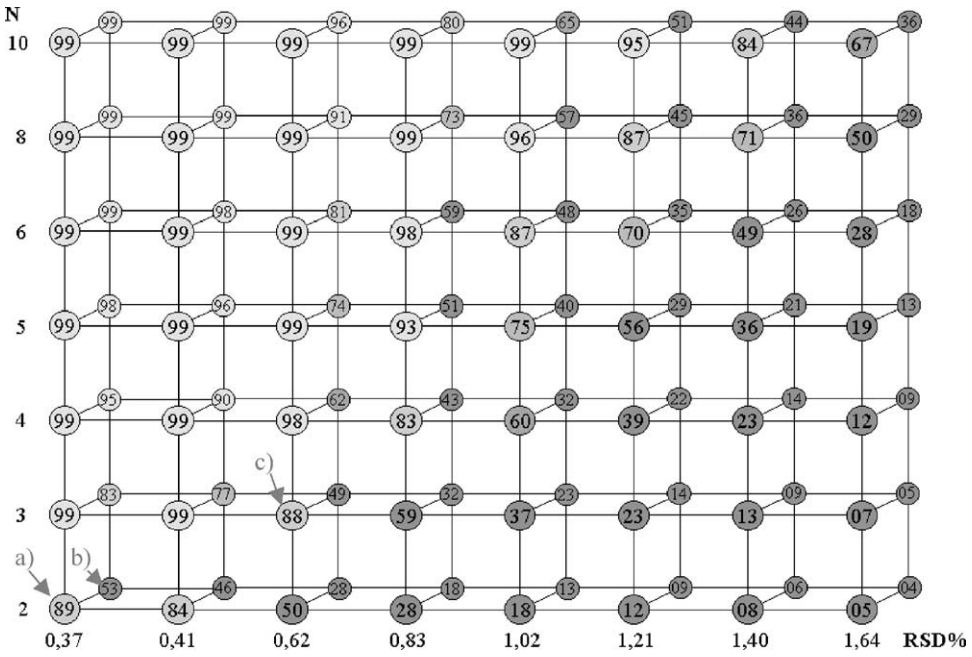

Fig. 3. The circles in this figure demonstrate the probability to correctly accept the method transfer, because the true bias (0% in the upper layer and 1% in the deeper layer) is lower than the 2% of the acceptance criteria. On the *x*-axis the variation R.S.D.% according to Eqs. (1) or (3) is depicted. The *y*-axis shows the number *N* of analysts or series used. The error probability α was 0.05. For explanation of (a), (b) and (c) see text.

following general equation (3):

$$\hat{\sigma}_{\bar{x}_{\text{series}}} = \sqrt{\hat{\sigma}_{\text{B}}^2 + \frac{\hat{\sigma}_{\text{SP}}^2}{n_{\text{SP}}} + \frac{\hat{\sigma}_{\text{SYS}}^2}{n_{\text{SP}} n_{\text{SYS}}}} \tag{3}$$

The $\hat{\sigma}_{\text{SYS}}$ and the $\hat{\sigma}_{\text{SP}}$ have only a small effect on the $\hat{\sigma}_{\bar{x}_{\text{series}}}$. Performing four executions ($n_{\text{SP}} = 12$) instead of three per sample under the same error structure of the starting example (see Section 2.1) leads to an only very slightly smaller variation of 0.36%:

$$\hat{\sigma}_{\bar{x}_{\text{series}}} = \sqrt{0.3^2 + \frac{0.6^2 + 0.3^2}{12}} = 0.36$$

Therefore just the value $\hat{\sigma}_{\text{B}}$ for the series means has been modified: 0.35, 0.58, 0.80, 1.00, 1.20, 1.40 and 1.60 have been taken into account. This results in R.S.D.-values for the variation of the series-means $\hat{\sigma}_{\bar{x}_{\text{series}}}$ of 0.41, 0.62, 0.83, 1.02, 1.21, 1.40 and 1.64. According to these extensions Fig. 3 demonstrates the effect of higher variations of the laboratories on the probability to correctly accept a method transfer. To retain a probability of more than 80%, which is regarded as adequate, for every 0.2%-step of the estimated total R.S.D.% at least one additional series would be necessary (individual R.S.D.%-values for the *x*-axis could be calculated by means of Eq. (3)). As one can see here, a fixed design cannot suffice in every case. For every special situation another individual number of series is required. At an R.S.D. of 1.02% and a number of series of 3 the probability decreases to 37%. A higher R.S.D. of the series-means needs a higher number of series, if one wants to identify a successful transfer with an adequate probability. Hence Fig. 3 provides various possible examples for a transfer situation. For different situations, where the individual error amounts of the involved error sources yield the same $\hat{\sigma}_{\bar{x}_{\text{series}}}$ according to Eq. (3), there is the same probability to accept the method transfer. Thus, with the knowledge of the own error sources and the total variation according to Eq. (3) and the number of involved series, Fig. 3 provides the probability for correctly accepting a successful transfer. Note, for different transfer situations, where the ratio of $\hat{\sigma}_{\bar{x}_{\text{series}}}$ to the acceptance criteria and the ratio of the bias to the acceptance criteria are the same, the probability to accept a transfer correctly is also equal. For e.g. the cases: (a) $\hat{\sigma}_{\bar{x}_{\text{series}}} = 0.8\%$, acceptance criterion = 2.0% and bias = 1.0% and (b) $\hat{\sigma}_{\bar{x}_{\text{series}}} = 1.0\%$, acceptance criterion = 2.5% and bias = 1.25% the same probability to correctly accept a transfer results.

5. Conclusions

The ISPE design provides adequate probabilities to accept successful method transfers correctly only for relatively small error amounts. For a total variation of more than 0.62% this fixed design leads to unsatisfying probabilities to reject successful transfers. For each definite error amount an individual number of series is needed. Thus Fig. 3 provides an overview for numerous error scenarios. To retain a power of more than 80%, which is regarded as adequate, for every 0.2%-step of the total R.S.D.% at least one additional series would be necessary.

Acknowledgements

We are very grateful to Dr. K. Baumann, Dr. J. Ermer, P. Hasemann, H. Köppel and Dr. B. Renger for critically reading the manuscripts and their numerous valuable comments and suggestions.

References

[1] ISPE Good Practice Guide for Technology Transfer available at http://www.ispe.org/.
[2] USP <1010> Analytical Data—Interpretation and Treatment, Pharmacopeial Forum 30 (2004) 236–263.
[3] D.L. Massart, B.G.M. Vandeginste, L.M.C. Buydens, S. De Jong, P.J. Lewi, J. Smeyers-Verbeke, Handbook of Chemometrics and Qualimetrics, Part A, Elsevier, Amsterdam, 1997, p. 88–91.
[4] C. Hartmann, J. Smeyers-Verbeke, W. Penninckx, Y. Vander Heyden, P. Vankeerberghen, D.L. Massart, Anal. Chem. 67 (1995) 4491–4499.
[5] U. Schepers, J. Ermer, L. Preu, H. Wätzig, J. Chromatogr. B 810 (2004) 111–118.
[6] MVA—Method Validation in Analytics, 2001, NOVIA GmbH, Frankfurt, Germany, available at http://www.novia.de/.