Available online at www.sciencedirect.com
SCIENCE DIRECT
Journal of Pharmaceutical and Biomedical Analysis 41 (2006) 290–292
JOURNAL OF PHARMACEUTICAL AND BIOMEDICAL ANALYSIS
www.elsevier.com/locate/jpba

Short communication

Application of the equivalence test for analytical method transfers: Testing precision using the United States Pharmacopoeia concept ⟨1 0 1 0⟩

U. Schepers, H. Wätzig *

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

Received 15 July 2005; received in revised form 21 October 2005; accepted 24 October 2005
Available online 9 December 2005

Abstract

In this work, the performance of the USP ⟨1 0 1 0⟩ concept for comparing precision has been investigated. A diagram has been constructed to relate common variance ratios, sample sizes and the corresponding powers. The choice of the upper acceptable limit of variance ratios strongly influences the power. Small upper limits, such as 2.25, are not practical. The proposed upper limit of 4 requires sample sizes of 14 or higher to achieve a power of 80%. If the precision of a method is very good, higher ratios seem to be acceptable, with a significant reduction in measurements. For example, using $n = 6$ is sufficient to obtain a power of 90%, if the variances are in fact the same and the acceptable variance ratio is 16.

Keywords: USP ⟨1 0 1 0⟩; Equivalence test; Analytical method transfer; Quality control; Pharmaceutical; Precision; Simulation; Acceptance criteria; Statistic

1. Introduction

The effective transfer of methods from one laboratory to another is a pre-requisite for sharing tasks and responsibilities within a company or within a network of companies that work together. General requirements for successful transfers are described in ref. [1]. In ref. [2], aspects of comparing mean values from different laboratories have been outlined, in relation to the ISPE guideline [3].

The well known *F*-test can be used focussing on the comparison of the variances of the participating laboratories [4]. However, the *F*-test principle to test on a statistical significance is connected with some drawbacks. For example, it is possible to detect a significance, which is of no practical importance if the estimated variance of the reference laboratory is very small and the variance of the receiving laboratory somewhat larger but acceptable.

The general chapter ⟨1 0 1 0⟩ of the USP proposes an alternative to compare the variances of two methods [5]. This approach could also be used in the context of analytical method transfers where one method is performed at two sites. In the USP approach, a confidence interval for the ratio of the (true) variances is calculated and afterwards compared to an a priori defined acceptable upper limit. The probability to wrongly accept a higher ratio of variances than predefined is advantageously controlled to 5% (=α). The other risk is linked to the probability to reject a transfer due to unacceptable precision though in fact it was acceptable. The probability to accept a method transfer correctly is the so called power. This power is related to the sample size, the chosen upper limit and the true, but unknown ratio of variances. These relationships are investigated in this article.

2. Test procedure

2.1. Method

The USP proposal tests whether or not the upper limit of the 90% confidence interval of the variance ratio of two methods exceeds an a priori defined ratio. Here, only the upper limit is of concern, as an improvement in precision is not problematical. The testing of the upper limit against the predefined ratio is constructed by means of the following Eq. (1):

$$\frac{\hat{\sigma}_2^2/\hat{\sigma}_1^2}{F_{\alpha,n_2-1,n_1-1}} \leq A \quad (1)$$

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

0731-7085/$ – see front matter © 2005 Elsevier B.V. All rights reserved.
doi:10.1016/j.jpba.2005.10.030

where $\hat{\sigma}_2^2$ is the estimated method variance of the receiving unit, $\hat{\sigma}_1^2$ the estimated reference method variance, A the predefined tolerated ratio of variances and F_{α,n_2-1,n_1-1} the upper α percentile of an F-distribution with $n-1$ numerator and $n-1$ denominator degrees of freedom. In ⟨1 0 1 0⟩, in an example α is suggested to be 0.05 and A to equal 4, respectively. So the upper 5% of the distribution are not included in the confidence interval.

Performing the test one compares the predefined acceptance value to the upper limit. One accepts the alternative hypothesis, H_1: $\theta < \theta_0$ of equivalent variances or a smaller variance of the receiving unit if the calculated term does not exceed the limit A. If the upper limit of the confidence interval exceeds the tolerated ratio one accepts the null hypothesis, H_0: $\theta \geq \theta_0$ and concludes that the both variances are not equivalent and the receiving unit variance is unacceptably high. For example, the variances of the sending unit and the receiving unit were estimated as 2.84 and 3.87, respectively (sample size $n = 12$ each). Then the upper 90%-confidence limit yield 3.89 (with $F_{0.05,11,11} = 0.35$), which is slightly lower than 4. In this case, the alternative hypothesis of at least equal variances cannot be rejected.

Note, that in principle this test needs two independent samples. The α-error cannot exactly be kept to the predefined 5% if the two samples are correlated to some extend. The total variation results from the variation of the method but also from the variation within the analyzed batch. A high contribution of the variation of the batch would cause a strong correlation between the results in the two labs. It is assumed here, that this contribution of the batch is not relevant (e.g. <10% of the total variance).

2.2. Calculation of the power

The power for a special test situation is accessible by an excel function given in ref. [5]: power = FDIST((*R*/*A*)FINV(α, n_2-1, n_1-1), n_2-1, n_1-1). R is the true ratio of variances at which the power is determined, A the maximum ratio of acceptance, α the significance level, typically 0.05 and n is the sample size. This function allows for the calculation of the power against a true (unknown) ratio of the participated variances. The resulting powers of several conceivable test situations have been calculated by this method.

3. Results, discussion and conclusion

A summary of several power calculations for different upper limits and sample sizes is given in Fig. 1. The choice of a higher upper limit strongly influences the power. A very small upper limit of 2.25 is not very practicable, as even with a sample size of 24 (which can be regarded as high) only a chance of 62% for the

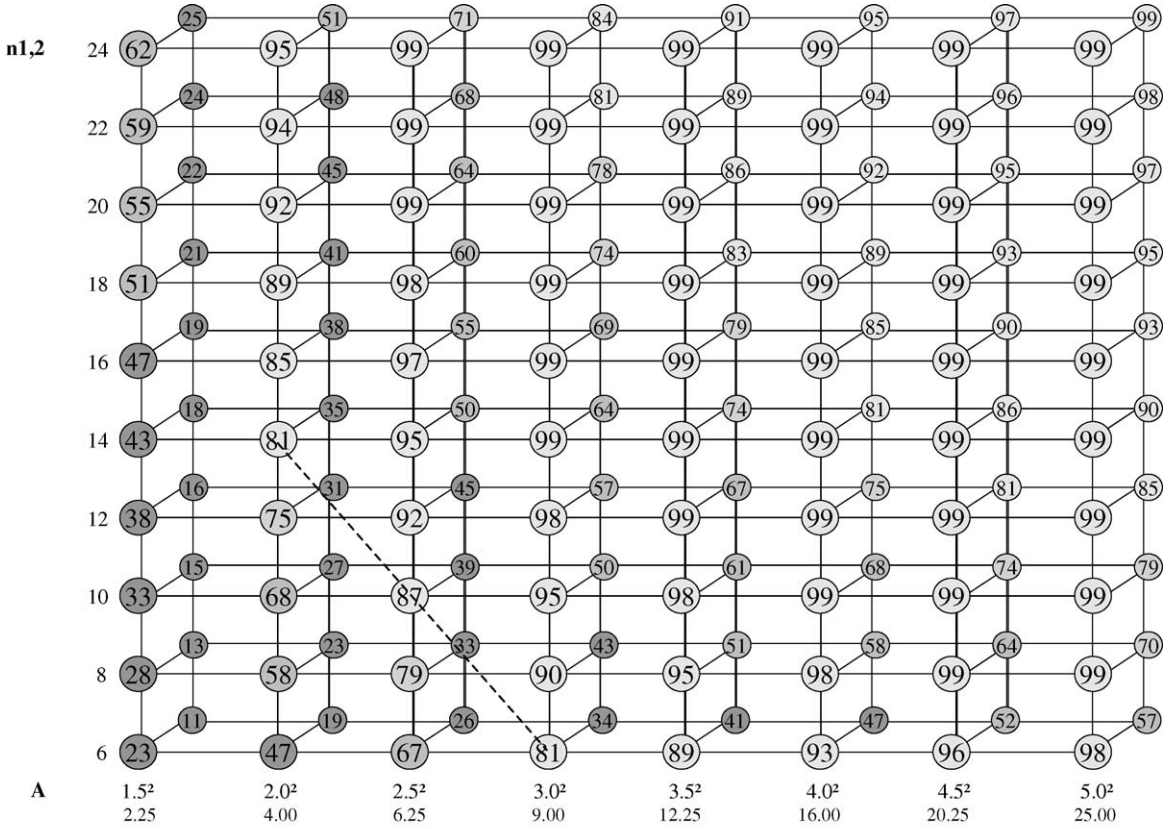

Fig. 1. The power of several scenarios is shown. On the x-axis, the predefined upper limit is given (in terms of variance ratios, depicted as squared ratios of standard deviations). The y-axis shows the sample size to determine the variances. The circles in the front layer contain the resulting power if in fact the variances are equal (resulting in a variance ratio of 1). For the layer of circles in the back the true ratio of variances R is calculated according to $[1+\{(\sqrt{A}-1)/2\}]^2$. This ratio is half-way between 1 (meaning equal variances) and the acceptable upper limit. So, for example, in the fourth column with an upper limit of 3.0^2 the true variance ratio of the back layer is: $[1+\{(\sqrt{3^2}-1)/2\}]^2 = [1+\{2/2\}]^2 = 2.0^2$ or 4.0.

acceptance exists, when in fact the variances are equal. The upper limit of 2.25 is also depicted as 1.5^2 to demonstrate that a ratio of standard deviations of 1.5 corresponds to a ratio of variances of 2.25. Furthermore, one can see, that an upper limit of more than 16 (corresponding to a ratio of 4 of the standard deviations) is smoothly to manage because only six samples are needed to correctly conclude to acceptance with a high probability. For example, such a high ratio of variances is regarded as acceptable, if the precision of the method is very good in the sending unit. Consider the corresponding RSD% was 0.2%: in this case a five-fold RSD% in the receiving unit would be well acceptable. Usually experienced analysts do not investigate on very small standard deviations, because they are aware of the rather high variability of standard deviations. A value of 0.2%, even when measured with a high data number, may be difficult to reproduce. On another day, even a minor additional error source which did not occur before could easily double such a small variability.

For the interval between 2^2 and 3^2 for every 0.25-step of the upper limit $(x+0.25)^2$ a decrease of the sample size of 2 can lead to retention of the power of more than 80%, starting with an upper limit of 2^2 and a sample size of 14 (dashed line in Fig. 1).

The above considerations are only valid if the true variances of the sending and receiving unit are in fact the same. However, it is also important to study, what power can be expected, if the true variances are different, but within the acceptable ratio.

If the true ratio R is just the acceptable ratio A, then the power is precisely 5%, according to the construction of the confidence interval. In order to understand, which power can be expected between these two extremes, we calculated the power for a half-way-ratio. This ratio is between 1 (meaning equal variances) and the acceptable upper limit A. It is calculated according to $[1+\{(\sqrt{A}-1)/2\}]^2$ and depicted in the back layer of circles in Fig. 1. For example, in the fourth column with an upper limit of 3.0^2 the true variance ratio R of the back layer is $[1+\{(\sqrt{3^2}-1)/2\}]^2=[1+\{2/2\}]^2=2.0^2$ or 4.0.

An adequate power of approximately 80% or above can only be achieved with sufficiently high data numbers. These numbers are noticeable higher than the ones needed to obtain sufficient power, if the variances are in fact the same. As can be derived from Fig. 1, an acceptable ratio A of 3^2 requires $n=22$. In order to reduce n, a more tolerant A must be chosen (e.g. $A=4^2$ requires $n=14$, $A=5^2$ corresponds to $n=10$).

To our understanding, values A, such as 16 $(=4^2)$ or above may often be the best choice. Consider $A=16$, $n=6$: the power drops strongly if one assumes a true bias. This also means, that the test will reject the method transfer in most cases, if the true ratio of the variances is between 7 and 16. The acceptance ratio of 16 does not mean at all, that it is likely to accept true ratios close to the value A. In addition, as outlined above, a generally low standard deviation may suggest more tolerant acceptance criteria.

Acknowledgements

We are grateful to PD Dr. Knut Baumann, Dr. Joachim Ermer, Heidi Köppel and Dr. Lutz Preu for critically reading manuscripts.

References

[1] J. Ermer, J.H.McB. Miller, Method Validation in Pharmaceutical Analysis, Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim (Germany), 2005, pp. 281–300.
[2] U. Schepers, H. Wätzig, J. Pharm. Biomed. Anal. 39 (2005) 310–314.
[3] ISPE Good Practice Guide for Technology Transfer available at www.ispe.org.
[4] 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. 104.
[5] USP 28-NF 23, General Chapter (1010), 2005.