Journal of Chromatography B, 877 (2009) 2270–2274

Contents lists available at ScienceDirect

Journal of Chromatography B

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

Short communication

A proposed "fixed" range decision criteria for transfer of bioanalytical methods☆

Kumar A. Shah, H. Thomas Karnes*

Department of Pharmaceutics, Virginia Commonwealth University, Richmond, VA 23298, USA

ARTICLE INFO

Article history:
Received 18 March 2009
Accepted 14 April 2009
Available online 17 April 2009

Keywords:
Bioanalytical method transfer
Acceptance criteria
Fixed range
Standard error of mean

ABSTRACT

Many approaches have been proposed for decision criteria to judge whether or not transfer of bioanalytical methods has been successful. Many of these approaches involve mathematical and statistical complexities that limit their use routinely. The FDA accuracy criterion (±15%) without an allowance for imprecision may be used for method transfer and may result in a large numbers of method transfers being judged unacceptable when the method is valid under both conditions. An acceptance criterion should be based on the existing guidance, be convenient and be based on statistical principles that provide consistent and reasonable rejection rates. In the current paper, we propose a "fixed" range acceptance criteria based on the FDA bioanalytical guidance limits on precision and accuracy. While the proposed "fixed" range criteria shares the shortcomings of any other fixed criterion, there are advantages when compared to use of accuracy criterion alone. The proposed criterion is also more user-friendly. Data simulations were performed to assess the probabilities of successful transfer using the proposed criteria. With an experimental design consisting of 3 independent runs with 3 replicates per run, a fixed criterion of ±20% of the reference method mean is proposed.

1. Introduction

Bioanalytical methods are routinely transferred and redeveloped for a number of reasons [1]. Transfer of a method requires assurance that the test laboratory or new method can obtain reliable and equivalent results. Comparability of between-study data is helpful to ensure proper interpretations of bioavailability, bioequivalence, pharmacokinetics, and toxicokinetics data. Cross validation and/or transfer of bioanalytical methods encompasses comparisons of data for two or more bioanalytical circumstances [1]. Interlaboratory and cross validation studies are generally evaluated using spiked matrix controls. At present there is no clear consensus on the most appropriate acceptance criteria or study design in such bioanalytical method data comparisons.

Several approaches have been described for the evaluation of method transfers [2–11]. Classical approaches used for evaluation of control data for bioanalytical method transfer include the independent validation approach, statistical difference testing and statistical equivalence testing. All these approaches have their own merits and drawbacks which have been extensively discussed [12]. Approaches based on a complete consideration of statistics are robust [4], but they may not be readily applicable in laboratories lacking knowledge in statistics. In these cases, it may be useful to investigate "fixed" criteria or rules that are based in part on statistics.

2. Proposed fixed criterion

Bioanalysts have been historically more prepared to accept criteria which are fixed and easy to use. The "4/6/15" rule is one such example in method validation. A criticism of this criterion is that it lacks a consistent statistical foundation. In the absence of an accepted guidance criterion for method transfer, many laboratories choose to use the fixed accuracy criteria of ±15% to compare the difference between means of the reference and the test method [7]. Since this involves a comparison of the means, it is logical to allow for random variation although no allowance is generally made. Failure to account for imprecision in such a comparison would lead to method transfers being judged unsuccessful, when the method is truly valid under both conditions. At the same time, failure to evaluate imprecision could also lead to the acceptability of method transfer when the transfer is truly not acceptable.

It would be desirable to employ approaches based on statistical considerations to establish fixed criteria rather than using fixed criteria based on a consensus opinion without statistical considerations. Such a proposal, like any other "fixed" criterion, may not be

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

* Corresponding author. Tel.: +1 804 828 3819; fax: +1 804 828 8359.
E-mail address: tom.karnes@vcu.edu (H.T. Karnes).

1570-0232/$ – see front matter © 2009 Elsevier B.V. All rights reserved.
doi:10.1016/j.jchromb.2009.04.020

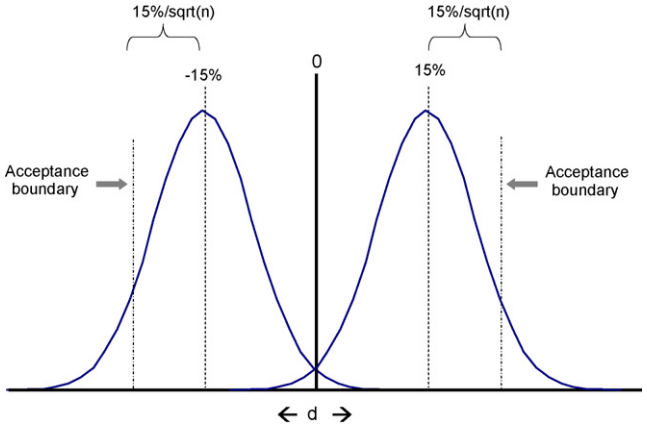

Fig. 1. Normal distributions (±15% CV) around the extremes of the FDA guidance accuracy criterion (±15%), the broadness of which is determined by the standard error of the means. The acceptable difference (d) between reference and test method is based on the maximum allowable error of the mean.

entirely statistically correct, but can have practical advantages over a straight comparison of means approach.

If μ_T is defined as the test method mean result, and μ_R as the reference method mean result, we propose that the acceptance criterion should be more accurately comprised of the sum of the FDA guidance accuracy limit of ±15% and the standard error of the mean of μ_T, the ratio of the standard deviation "s" of the test method to the square root of the number of replicate measurements 'n'. The standard error of the mean provides a gauge for how variable the mean can be expected to be when performing n replicate analyses [13]. Considering the FDA guidance on precision of 15% CV [1], fixed acceptance limits for bioanalytical method transfer based on the FDA guidance acceptance limit maxima can be set as $\pm(15\% + 15\%/\sqrt{n})$. Fig. 1 depicts the proposed acceptance criterion that takes into account method imprecision. In this context, the x-axis shows the estimated mean difference (d) between the two methods. A risk analysis can then be employed to determine the probability that the test laboratory would observe results outside the stated acceptance limits of $\pm(15\% + 15\%/\sqrt{n})$. This provides a tool to determine a reasonable fixed criterion for bioanalytical method transfer. Such a tool would depend on the number of samples, the number of times the samples are run, and other experimental details such as whether or not the samples are paired as in the case of clinical samples or unpaired in the case of quality control samples.

3. Experimental

Computer simulations were performed using Crystal Ball (ver. 7.3.1) software in order to evaluate the probability of successful transfers for different experimental designs. The probability of success for the proposed "fixed" range approach was compared with the FDA prescribed accuracy criterion without accounting for precision. A one-way random effects statistical model described below was used to generate data for the reference and test laboratories:

$$Y_{ij}^{R} = \mu + \varepsilon_B^R + \varepsilon_W^R \quad \text{and} \quad Y_{ij}^{T} = \mu^T + \varepsilon_B^T + \varepsilon_W^T$$

where Y_{ij} is the jth (J = 1, 2, …, J) replicate observation from the ith (I = 1, 2, …, I) run, μ is the true analytical mean, ε_B and ε_W are the errors associated with between- and within-run variability. The super-scripts R and T represent the reference and the test laboratories, respectively. The random errors ε_B and ε_W are assumed to be normally and independently distributed with means zero, and variances σ_B^2 and σ_W^2, respectively. The total analytical variance $\sigma_{TOT}^2 = \sigma_B^2 + \sigma_W^2$. The data were simulated in a manner so that the total analytical variance of the reference method does not exceed the FDA criterion of 15% CV.

Table 1
Parameters used for simulations[a,b].

Variance component	Estimate
Concentration (ng/mL)	0.5, 4, 80
Within-run %RSD	4, 7, 10
Variance ratio (R) = σ_B^2/σ_W^2	0.5, 1‘
Bias (%) between the averages of reference and test laboratories	14, 15, 16, 18, 20, 22, 25, 30

[a] The number of replicates and runs was equal for the reference and the test method.
[b] 10,000 simulations were run in each case.

The variance of the reference and the test method were assumed to be the same for the various simulations performed. Parameters used for data simulation are depicted in Table 1. Various combinations of experimental designs were simulated. The number of replicates simulated were J = 3 and 6, and the number of runs simulated were I = 3 and 6. In each case, 10000 simulations were run. The limits of acceptance based on the proposed fixed range approach were set as $\pm(15\% + 15\%/\sqrt{n})$ where n is the number of samples evaluated. The probability of a successful transfer for the proposed fixed range acceptance criterion was then compared with the ±15% difference between the means criterion without any allowance for imprecision.

4. Results and discussion

Fig. 2 shows the probability of successful transfer for the test method based on the two separate decision criteria for simulated results involving various experimental scenarios. As seen in this figure and as expected, there is a clear difference in the probabilities of a successful transfer for the two criteria. For example, looking at Fig. 2A, when the true relative bias between the means of the two methods is 15%, considering the ±15% criterion without accounting for imprecision, the probability of rejecting a transfer is as high as 50% for all combinations of runs and replicates. Such a large probability of rejection may not be desirable, especially if the methods are "equivalent" considering a combined estimate of trueness and imprecision. Conversely, if the limits of acceptance are established to be $\pm(15\% + 15\%/\sqrt{n})$ we can conclude from the simulation that the probability of an unsuccessful transfer is less than or equal to 5% for all experimental designs. Combining the estimates of systematic error and its uncertainty will therefore save the time, labor and cost of generating unnecessary data.

If one compares the situations represented by Fig. 2A and B, it can be observed that for both the simulated criteria, methods that are more precise are better rewarded. In other words, when the relative bias is within the acceptance criterion, the probability of accepting a transfer is greater with more precise methods. Conversely, when the relative bias is outside the acceptance criteria, the probability of rejecting a transfer increases with more precise methods. There however appears to be a better correlation for the proposed fixed range acceptance criteria in this regard as seen by the steeper slopes of the probabilities of a successful transfer for simulations with a 4% RSD as compared to simulations with a 10% RSD. Thus, as the precision of a method improves, the chance that a correct decision is made is better controlled with the $\pm(15\% + 15\%/\sqrt{n})$ criteria as compared to the ±15% criteria.

Another interesting observation is seen when the true bias between the means of the two methods is exactly at the boundary of the FDA acceptance limit for accuracy (±15%) [1]. Looking at Fig. 2A, a node is observed when one views the probability of successful transfer for various experimental designs. This indicates that at the boundary of acceptance, the probability of a successful trans-

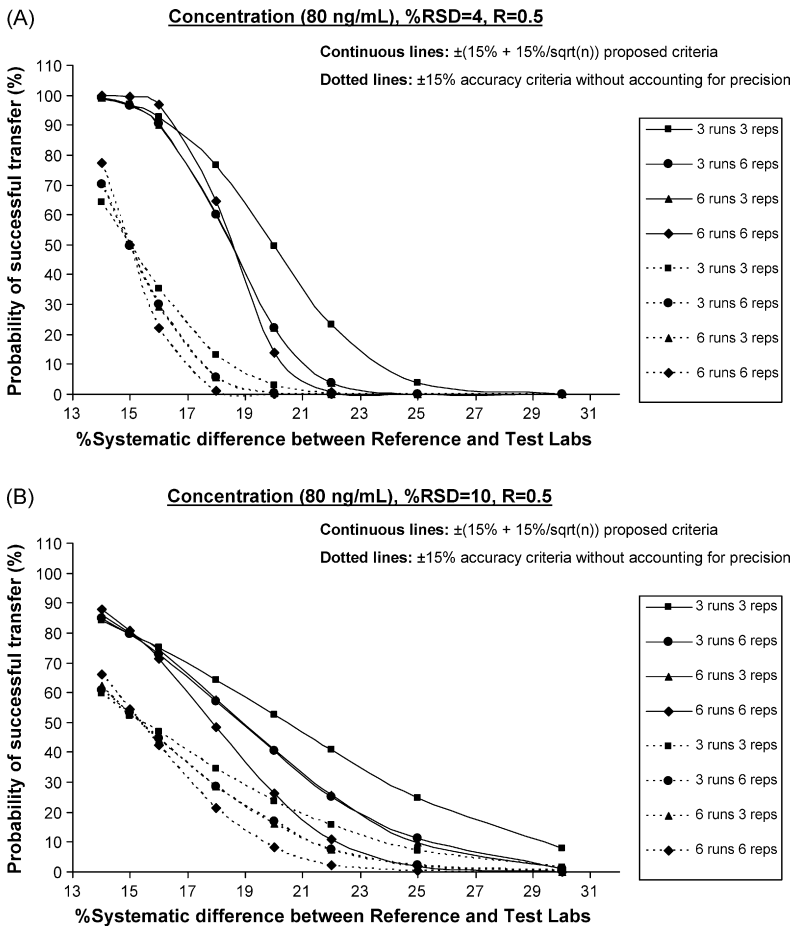

Fig. 2. Probability of successful transfer as a function of % relative bias between test and reference methods for simulated results and various experimental scenarios. (A) Conc. = 80 ng/mL, within-run RSD = 4%, $R = 0.5$, (B) Conc. = 80 ng/mL, within-run RSD = 10%, $R = 0.5$.

fer does not change even as the experiments become more rigorous. The failure rate of a 3 replicate 3 run design is about 50% and is the same for a 6 replicate 6 run simulation. Such an outcome punishes rigorous experimental designs in their ability to accept transfers. No such node is observed in the case of the proposed fixed criteria. In fact, the acceptance limits changes in this case as the experimental design changes. For instance, the probability of accepting a transfer with a 3 run 3 replicate design is about 50% at the acceptance boundary of ±20% [i.e. $\pm(15\% + 15\%/\sqrt{9})$]; however, the probability of accepting a transfer increases to approximately 75% with a 6 run 6 replicate design at the acceptance boundary of 17.5% [i.e. $\pm(15\% + 15\%/\sqrt{36})$] as observed in Fig. 2A. The, simulation results also revealed that the probability of accepting a comparison is constant over a broad range of concentrations for varying values of R (i.e. σ_B^2/σ_W^2) using both criteria for the various experimental designs (results not shown).

Based on the simulations, the simplest experimental design of 3 runs and 3 replicates provided reasonable results using the proposed fixed range criterion. Experimental designs that employ a greater number of runs and/or replicates provide an increased probability of success for systematic error differences of 15% or less. It is more throughput efficient, however, to use experimental design with fewer replicates and runs in order to detect differences. Similarly, for systemic errors outside the acceptance range, the results can be made more reliable with an increase in the number of replicates and runs, resulting in an increased rejection rate.

Based on a 3 replicates and 3 runs experimental design, a fixed range acceptance criterion can be proposed as $\pm(15\% + 15\%/\sqrt{9}) = \pm 20\%$. Since the criterion is derived based on a consideration of one standard deviation, the probability that the true mean would be expected to be within the defined range would be 67%. Each concentration would require individual evaluation must conform to the acceptance criterion for a method to be considered equivalent. A 3 run, 3 replicate design would thus result in a consistent, reliable and easy to remember rule of thumb which we will refer to as the 3/3/20 rule based on the number of runs (3), the number of replicates (3) with a maximum allowable relative bias of the mean test result of ±20%. Similarly, for a 3 run 6 replicate experimental design, the fixed range acceptance criteria can be proposed as $\pm(15\% + 15\%/\sqrt{18}) = \pm 18.5\%$ (=18% approximately) of the reference method mean. A 3/6/18 rule can be suggested in this case resulting in a more rigorous experimental design and an improved ability to detect differences at the expense of requiring a larger number of experiments, however.

Although the proposed acceptance criteria controls the risk of α errors more efficiently as compared to the ±15% criterion without an allowance for precision, it increases the risk of committing a β error. This is evident from the results of simulations depicted in Fig. 2. The proposed criterion is a "fixed range criterion" however, and a drawback of any fixed criteria is that it is not possible to simultaneously control both α and β errors. Certain laboratories might want tighter control of β errors. Rigorous statistical approaches are acknowledged to be more correct in addressing this problem. However, fixed criteria are universally used and accepted in regulated environments. In order to address the issue of β errors, an alternative approach is discussed further.

An alternative approach to limit β error, would be to reduce the accuracy criteria of ±15% prescribed by the FDA by a factor

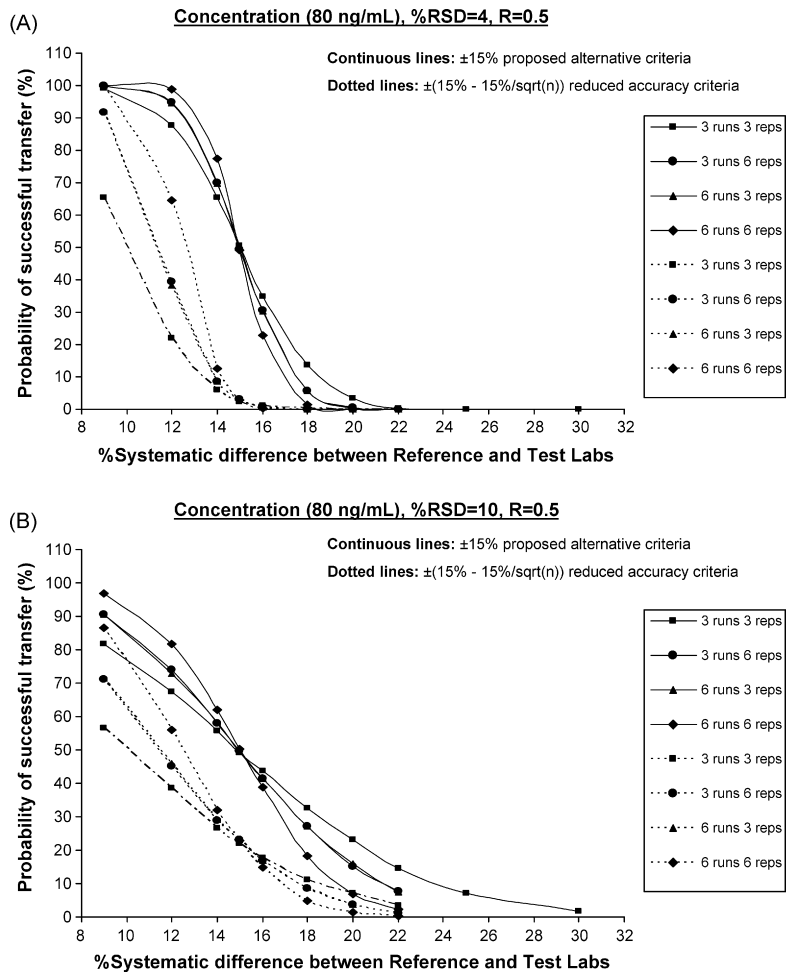

Fig. 3. Probability of successful transfer using a more conservative accuracy criterion as a function of % relative bias between test and reference methods for simulated results and various experimental scenarios. (A) Conc. = 80 ng/mL, within-run RSD = 4%, $R = 0.5$ (B) Conc. = 80 ng/mL, within-run RSD = 10%, $R = 0.5$.

equal to $15\%/\sqrt{n}$. In this substitute approach, the proposed fixed criterion would then be a maximum allowable acceptance limit of ±15% and would still account for imprecision. The proposed acceptance limit is a sum of the more conservative accuracy criterion of $\pm(15\% - 15\%/\sqrt{n})$ and the maximum allowable standard error of the mean of the test method i.e. $15\%/\sqrt{n}$. This would result in a maximum acceptance limit of ±15% for method comparison that would depend on the experimental design. Fig. 3 shows representative simulation results for the modified criterion. The parameters simulated were similar to those previously described. The systematic bias simulated was in the range of 9–30%. It can be observed from these results that the probabilities have been reversed when compared with the earlier described $\pm(15\% + 15\%/\sqrt{n})$ criterion. For example, when the proposed criterion was fixed at $\pm(15\% + 15\%/\sqrt{n})$, the risk of making a false negative error for a 6 run 6 replicate design at the acceptance boundary of 17.5% was approximately 70%. This is evident from Fig. 2A. This risk of committing a false negative error at the boundary of acceptance, however, reduces to about 50% for the same experimental design when the acceptance criterion is fixed at ±15% with the more conservative accuracy criterion, as seen in Fig. 3A. As in the previous case, Fig. 3A and B show that more precise methods have better control over falsely rejecting truly comparable methods as well as falsely accepting truly non-equivalent methods. Simulation outcomes based solely on the accuracy criterion (shown by dotted lines in Fig. 3A and B) would result in a greater incidence of rejecting truly comparable methods.

5. Conclusion

Several approaches have been suggested in the literature to address the issue of analytical method transfer and method comparisons. The statistically rigorous total error approach [4] is perhaps the most robust and scientifically correct. However, this may require adequate knowledge of statistics as well and an evaluation of the variability of the methods being tested. A ±15% difference between the means criteria, without allowance for precision, does not provide a reasonable approach to address the risk of false conclusions regarding method comparisons. In the current paper, we have proposed a user-friendly “fixed” range acceptance criteria based on a consideration of method accuracy and precision. We have combined a consideration of the maximum allowable FDA bioanalytical guidance limits on precision and accuracy and the ease of application of a fixed range approach. Various experimental designs were evaluated and although there are many options, an experiment with 3 runs with 3 replicates each at 3 concentrations with an acceptance range of ±20% is proposed. This general approach could be applied to individual dosed subject samples or prepared controls. We believe the 3/3/20 criterion for method comparisons to be reasonable in terms of its ease of use, the number of required experiments and the control of false positive results. In instances where a greater control over false negative results is required, a reduced accuracy based fixed criterion is suggested. The proposed fixed range approach could also be extended to many other validation situations in which data comparisons are involved.

For example, in the case of incurred sample reanalysis, an experimental design could be anticipated that would involve comparing observations with $n = 1$ for each comparison. An acceptance criterion consistent with this approach and the established guidance would then be $\pm(15\% + 15\%/\sqrt{1})$ or $\pm 30\%$.

References

[1] http://www.fda.gov/CDER/GUIDANCE/4252fnl.htm – Guidance for Industry Bioanalytical Method Validation.
[2] http://www.ispe.org – ISPE good practice guide for technology transfer.
[3] U. Schepers, H. Watzig, J. Pharm. Biomed. Anal. 39 (2005) 310.
[4] W. Dewé, B. Govaerts, B. Boulanger, E. Rozet, P. Chiap, Ph. Hubert, Chemom. Intell. Lab. Syst. 85 (2007) 262.
[5] C. Hartmann, J. Smeyers-Verbeke, W. Penninckx, Y. Vander Heyden, P. Vankeerberghen, D.L. Massart, Anal. Chem 67 (1995) 4491.
[6] D. Chambers, G. Kelly, G. Limentani, A. Lister, R. Lung, E. Warner, Pharm. Technol. 9 (2005) 64.
[7] A.J. DeStefano, Issues In The Transfer Of Methods To Analysts And Laboratories, in: AAPS Third Bioanalytical Workshop, Arlington, VA, May 1–3, 2006.
[8] U. Schepers, H. Watzig, J. Pharm. Biomed. Anal. 41 (2006) 290.
[9] S. Feng, Q. Liang, R.D. Kinser, K. Newland, R. Guilbaud, Anal. Bioanal. Chem. 385 (2006) 975.
[10] J. Vial, A. Jardy, P. Anger, A. Brun, J. Menet, J. Chromatogr. A 815 (1998) 173.
[11] G. de Fontenay, J. Pharm. Biomed. Anal. 46 (2008) 104.
[12] R. Kringle, R. Khan-Malek, F. Snikeris, P. Munden, C. Agut, M. Bauer, Drug Inf. J. 35 (2001) 1271.
[13] D.L. Massart, B.G.M. Vandeginste, S.N. Deming, Y. Michotte, L. Kaufman (Eds.), Chemometrics: A Textbook., Elsevier, New York, 1988, p. 31.