DISCUSSION FORUM

Accred Qual Assur 3:497

Hassan Y. Aboul-Enein

Some considerations in the use of internal standards in analytical method development

Key words Internal standard · Uncertainties · Analytical method developments

A number of analytical methods require the use of internal or external standards. However, there is no set of rules that can help the analyst make a choice.

An internal standard is usually chosen because it behaves in a similar way to the analyte under investigation. Usually, internal standards are added before sample pre-treatment steps, so that any losses of the analyte during sample preparation can be paralleled by losses of the internal standard. Accordingly, the standard will correct for these losses.

However, the following problems are encountered with the use of internal standards:

1. It is preferable to choose an internal standard which has a similar chemical structure to the analyte in order to mimic its behaviour. However, the behaviour of the internal standard in evaluating sample loss accurately is often overlooked.
2. During addition of the internal standard, measurement errors due to weighing, dilution and/or dispensing might occur which contribute to the total uncertainty of the system, which can be avoided by the use of an external standard. The effects of these errors should be taken into consideration especially in trace analysis as their impact could be essential.
3. Addition of an internal standard requires additional precautions to ensure that no interferences are present to compromise the measurement of the analyte. These interferences become serious especially during analysis of the analyte in a complex matrix such as biological fluids.

Therefore, an analytical method should be tested with and without an internal standard in order to check if there are advantages to warrant its use, to provide better validated results.

H.Y. Aboul-Enein
Biological and Medical Research Department (MBC-03),
King Faisal Specialist Hospital and Research Centre,
P.O. Box 3354, Riyadh 11211
Saudi Arabia

DISCUSSION FORUM

New ideas, controversial or even contentious papers – Please give us your views
Papers published in this section do not necessarily reflect the opinion of the Editors and the Editorial Board

Accred Qual Assur 3:497–498

Robert George Visser
Wim Oussoren

Remarks on Linsinger TPJ, Kandler W, Krska R, Grasserbauer M

The influence of different evaluation techniques on the results of interlaboratory comparisons

Accred Qual Assur (1998) 3:322-327

Introduction

In this journal, Linsinger et al. recently discussed the influence of different evaluation techniques on the results of interlaboratory comparisons [1]. They compared several outlier tests and concluded that Hampel's test due to its strictness is most suitable, both for the estimation of the 'true value' and for estimation of the 'true standard deviation'. Thus the use of Hample's test will give the best evaluation of laboratory performance using z-scores. In this article we will present an even better strategy.

Linsinger et al. also compared three different summarising indices. It was concluded that the calculation of the investigated indices proved to be of limited usefulness for the evaluation of overall laboratory performance and that only parameters which are relatively insensitive to an outlier test can be used. We only agree with the last part of this conclusion, but not with the first part. We have also evaluated both of the above points: the influence of outlier elimination on the standard deviation and the choice of summary index. Based on our conclusions we have chosen totally different solutions to the problem.

Background

Proficiency studies are very useful for the improvement of comparability of test results of analytical laboratories. This is now recognised by most analytical laboratories in many fields and participation in proficiency tests is increasing. Many organisers of proficiency tests for analytical laboratories use protocols based on standardised methods such as the well-known ISO 5725 [2]. The performance of the participating laboratories is often evaluated by means of z-scores but the problems and pitfalls that are involved with this strategy are well known and have been discussed in numerous articles. Each professional organiser of proficiency tests chooses his own solution to these problems. Some organisers use robust statistics (e.g. ASTM E2, EPA and Quasimeme), while others try to optimise the procedures using traditional statistics. Linsinger et al. [1] have discussed only the last option.

Estimation of the 'true value'

The estimation of the 'true value' of a parameter is essential in the evaluation of a proficiency test. To guarantee a close resemblance of the samples that the participating laboratories routinely analyse, many organisers of proficiency tests use natural matrix materials. The problem with 'real world' samples (e.g. we use gasoline, diesel, kerosene, crude oil, fuel oil) is that the 'true values', in principle, are unknown. Thus the 'true values' are usually estimated via calculation of some kind of (trimmed) mean. Some organisers use robust statistics in order to tackle the problems of influence of outliers and of non-Gaussian distribution. However, robust statistics are not yet described as a standard method and are also not accepted nor well understood by participating laboratories. Therefore, we have chosen to use the procedure of ISO 5725 and use traditional statistics and iterative use of Dixon's and Grubb's outliers tests to calculate a (trimmed) mean value. In most cases, when the number of data is large ($n>20$) and the distribution is (near to) Gaussian, the calculated mean is equal to a robust estimator like Tukey's biweight mean [3].

Calculation of the standard deviation

The estimation of the standard deviation of a parameter is necessary for the performance evaluation of the laboratories participating in a proficiency test. The problem with the standard deviation calculated from the results of the participating laboratories is that this standard deviation is dependent on many random factors: the sample matrix, the quality of all participating laboratories, the detection and rejection of outliers and therefore the estimation of the 'true value' and the choice of outlier test. Some organisers use robust statistics to calculate the spread (e.g. calculation of the DoD [4]) in order to tackle (some of) these problems. However, again the argument that robust statistics are not yet described in a standard method and also not accepted or well understood by participating laboratories, induced us to choose another strategy. Instead of a standard deviation calculated from the data of the participants, we calculate a 'target standard deviation' from a reproducibility published in the literature (e.g. in standards like ASTM, ISO or EN). This value may be fixed or concentration dependent, but in any case this 'target value' is independent of the above mentioned random factors. Through this independence an extra advantage is present: the results of two subsequent proficiency tests become fully comparable and performance changes in time are made easily visible. In case no literature values are available, other targets may be chosen depending on the goal of the proficiency test or the importance of the test. Any realistic value can be used, but we prefer to use a value based on the Horwitz equation [5].

Calculation and use of summary indices

The evaluation of proficiency tests with more than one parameter is not only of interest to the laboratory technicians, but also to laboratory managers who want to know how their laboratories have scored. Managers do not have time to study a technical report extensively and they like to have just one mark for the laboratory performance just as one gets in high school. And although one mark as a performance indicator may have its limitations, half a loaf is better than no bread.

In the literature, several procedures for calculation of summary indices are described [6]. Linsinger et al. concluded that there is little correlation between the three methods they compared. We agree with that conclusion, but we disagree with the explanation given. The observed lack of correlation is merely a result of the fact that not-comparable, mathematically different procedures are compared. For example, in the calculation of the RSZ-score negative z-scores and positive z-scores compensate each other, while in the calculation of the average of absolute z-scores these scores enlarge each other.

Rejecting all these and other methods described by Thompson and Wood [6], we have gone for a different approach although still based on z-scores. The resulting summary indices have satisfied us and our participants so far. Linsinger et al. advise that one should not use summary indices at all, but in view of the above, we do not agree with this advice.

Conclusion

In spite of the problems and in contradiction with the conclusion of Linsinger et al., we believe that it is possible to evaluate the performance of laboratories via z-scores and traditional statistics in a consistent and sound manner, and to use summarising indices as well. Our procedure may be not ideal, but we see it as a fairly good start. One cannot expect to develop the ultimate summary index immediately. After use of it in practice and after discussions with participants and other organisers of proficiency tests it may lead to a method that could be standardised. Other problems (e.g. n is too small or distributions are not Gaussian) are more dangerous and harder to tackle.

References

1. Linsinger TPJ, Kandler W, Krska R, Grasserbauer M (1998) Accred Qual Assur 3:322–327
2. ISO 5725 (1994) Accuracy (trueness and precision) of measurement methods and results. ISO, Geneva, Switzerland
3. Kafadar K (1982) J Am Stat Assoc 77:416–424
4. Beyrich W, Golly W, Peter N, Seifert R (1990) KfK 4721, EUR 11398 EN, PWA 15/90, Kernforschungszentrum Karlsruhe
5. Thompson M, Lowthian PJ (1997) J AOAC Int 80:676–679
6. ISO/IEC Guide 43 (1996) Proficiency testing by interlaboratory comparisons. ISO, Geneva, Switzerland
7. Thompson M, Wood R (1993) J AOAC Int 76:926–940

R.G. Visser (✉) · W. Oussoren
Institute for Interlaboratory Studies
P.O. Box 8204,
3301 CE Dordrecht, The Netherlands
Tel.: +31-78-654 5362
Fax: +31-78-651 2704