Talanta 82 (2010) 1995–1998

Contents lists available at ScienceDirect

Talanta

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

Short communication

Intra-laboratory assessment of method accuracy (trueness and precision) by using validation standards

A. Gustavo González*, M. Ángeles Herrador, Agustín G. Asuero

Department of Analytical Chemistry, University of Seville, 41012 Seville, Spain

ARTICLE INFO

Article history:
Received 27 June 2009
Received in revised form 19 July 2010
Accepted 27 July 2010
Available online 4 August 2010

Keywords:
Accuracy
Trueness
Precision
Validation standards

ABSTRACT

Assessment of accuracy of analytical methods is a fundamental stage in method validation. The use of validation standards enables the assessment of both trueness and precision of analytical methods at the same time. Procedures of intra-laboratory testing of method accuracy using validation standards are outlined and discussed.

1. Introduction

The accuracy of an analytical method is a primarily role for validation purposes [1–6]. In a former article [7] the assessment of method accuracy from recovery assays based on spiked matrices and spiked samples was revised. However in the mentioned paper, accuracy was considered in the perspective of a systematic component of the error, today superseded by the term 'trueness'. In his modern sense, accuracy is a performance characteristic that refers to the total error (systematic and random errors) and comprises two components: Trueness and precision, whose quantitative expressions are the 'bias' and the 'standard deviation' respectively [8,9]. These two figures of merit can be assessed independently, but it is possible to assess accuracy in a holistic way according to the measurement uncertainty and accuracy profiles [10–13]. In-house or intra-laboratory assessment of trueness and precision can be performed at a time when validation standards (VS) are available. VS have to be prepared in the same matrix as the expected for future samples. Certified or internal reference materials represent the best way to obtain VS, but spiked samples can be considered as a suitable alternative [14–16]. In the case of pharmaceutical formulations or other manufactured products where a 'placebo' is available, the bias or precision study can be carried out using spiked placebos. But, when the placebo is not available, selected stable samples fortified to a suitable level of analyte may be prepared. VS must be stable, homogeneous and as similar as possible to the future samples to be analyzed and they represent, at the validation stage, the future samples that the analytical procedure will have to quantify. Each VS have to be prepared and treated independently as a future sample. This independence is essential for a good estimation of the between-conditions variance in the assessment of accuracy; other important performance characteristics are assumed to be consistent with the fitness for purpose of the analytical method. Thus, it will be supposed that the method is suitably selective and sensitive and the possible matrix effects have been previously studied and corrected at the calibration stage [5,6]. In such a case, a corrected inverse prediction equation will be available to transform the measured analytical signal coming from the sample into the analyte concentration. Once these requirements meet, the test of method accuracy can be carried out. The aim of the present paper is to outline and discuss the most suitable and practical procedures for assessing the trueness and precision of an analytical method when VS are available, either independently or in a global way through the measurement uncertainty and the accuracy profiles.

2. Assessment of trueness and precision from a nested design and ANOVA calculations

According to ICH guidelines [1], precision may be considered at three levels: repeatability, intermediate precision and reproducibility. Repeatability expresses the precision evaluated under the same experimental conditions over a short time interval. Sometimes it is termed as intra-assay or within-run precision and refers to the "pure" random error associated with the assay measurement

* Corresponding author. Tel.: +34 954557173; fax: +34 954557168.
E-mail address: agonzale@us.es (A.G. González).

0039-9140/$ – see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.talanta.2010.07.071

process. Intermediate precision applies to within-laboratory variations: different days, different analysts, different equipment and so forth. Intermediate precision is sometimes called between-run or inter-assay precision. But as Peters and Maurer pointed out [17], *stricto senso*, intermediate precision is the total precision under varied conditions that it is expected within laboratory in a future assay, whereas so called inter-assay, between-run or between-day precision only measure the precision components caused by the respective factors. Nevertheless, if a between-day precision study is performed by spacing out the measurement days in such a way that other items (analysts, equipment, stock solutions, glassware, etc.) really changed, then this precision measurement could be considered as a truly intermediate precision estimation. Reproducibility, in its turn, expresses the between-laboratories precision like in collaborative studies. Reproducibility only has to be studied, if a method is supposed to be used in different laboratories. Unfortunately, some authors use the term "reproducibility" for intra-laboratory precision studies at the level of intermediate precision [18]. On the other hand, the trueness of an analytical procedure expresses the closeness of agreement between the mean value obtained from a series of measurements and the value which is accepted either a conventional value or an accepted reference value like VS. Trueness can be expressed in terms of recovery or absolute or relative bias [11]. Both trueness and intermediate precision studies can be carried out by predicting the actual concentrations of a series of VS selected. Following the golden rules of method validation [19], the analytical procedure should be validated separately for each kind of matrix considered as a whole (including sample treatments prior to analysis) and covering the full range of analyte concentrations. Accordingly it is advisable to perform the accuracy study with VSs at least for three concentration levels m (low, medium and high) covering the dynamic working range previously established in the calibration stage, with a number of n replicates at each concentration. The ICH Q2(R1) document [1] recommends three replicates and the FDA document [2] consider five replications. Accordingly, 3–5 replications are advisable. Calculations of intermediate precision and bias have to be carried out on results instead of analytical responses. Considering the different p days as the main source of variation for the intermediate precision study, a one-way analysis of variance (ANOVA) can be performed for each VS. Thus, for each concentration level m, the predicted concentration of the VS (by using the suitable calibration curve) will be x_{ij} with two indices: i (from 1 to p) corresponding to the different days and j (from 1 to n) accounting for the replications. From the ANOVA the estimations of within-days variance (S_W^2) and between-days variance (S_B^2) are obtained [5,20,21]. The within-days variance is also known as repeatability variance (S_r^2) and is given by

$$S_r^2 = S_W^2 = \frac{\sum_{i=1}^{p}\sum_{j=1}^{n}(x_{ij} - \bar{x}_i)^2}{p(n-1)} \quad (1)$$

with

$$\bar{x}_i = \frac{\sum_{j=1}^{n} x_{ij}}{n}$$

The between-days variance is calculated from

$$S_B^2 = \frac{\sum_{i=1}^{p}(\bar{x}_i - \bar{\bar{x}})^2}{p-1} - \frac{S_r^2}{n} \quad (2)$$

with

$$\bar{\bar{x}} = \frac{\sum_{i=1}^{p}\sum_{j=1}^{n} x_{ij}}{pn}$$

Table 1
Acceptable *RSD* values obtained from the Horwitz function and from the AOAC Peer Verified Methods program according to the concentration level of analyte (ppm = parts per millions, ppb = parts per billions).

Analyte (%)	Analyte fraction	Concentration unit	% RSD_H	% RSD_{AOAC}
100	1	100%	2	1.3
10	10^{-1}	10%	2.8	1.8
1	10^{-2}	1%	4	2.7
0.1	10^{-3}	0.1%	5.7	3.7
0.01	10^{-4}	100 ppm	8	5.3
0.001	10^{-5}	10 ppm	11.3	7.3
0.0001	10^{-6}	1 ppm	16	11
0.00001	10^{-7}	100 ppb	22.6	15
0.000001	10^{-8}	10 ppb	32	21
0.0000001	10^{-9}	1 ppb	45.3	30

The intermediate precision variance (S_{IP}^2) can be estimated according to [5]:

$$S_{IP}^2 = S_r^2 + S_B^2 \quad (3)$$

From this value, the corresponding relative value, RSD_{IP} is computed and can be compared with the expected values issued from the Horwitz equation and the "Horrat" parameter [22,23]. Horwitz [24] found an expression for predicting the expected value of the relative standard deviation for inter-laboratory trials according to:

$$\%RSD_H = 2^{(1-0.5\log C)} \quad (4)$$

where C is the analyte concentration expressed in decimal fraction. The RSD_H value is a primary criterion for evaluating reproducibility precision [25]. Intermediate precision predicted *RSD* is approximately one-half to two-thirds the RSD_H Horwitz value [26,27]. The "Horrat" value is often used as a benchmark for the performance of analytical methods, which is defined as the ratio of the actual relative standard deviation *RSD* calculated from the analytical data to the predicted Horwitz value:

$$Horrat = \frac{RSD}{RSD_H} \quad (5)$$

In our case, for evaluating the intermediate precision we take $RSD = RSD_{IP}$ in Eq. (5). The acceptable Horrat value for intermediate precision studies should not be higher than 1.3 [26].

Aside from Horwitz's parameters, values of *RSD* according to the AOAC Peer verified Methods Program [28] can be also considered. Both Horwitz and AOAC acceptable *RSD* values as a function of analyte concentration are presented in Table 1. As a quick rule, our RSD_{IP} result should be compared with one-half the corresponding *RSD* value appearing in Table 1.

The assessment of trueness can be performed according the same ANOVA results. Accordingly trueness can be expressed as the bias or the recovery obtained for each VS assayed [29].

In the bias calculation, the total mean value x is taken as the final result corresponding to the concentration of the VS, whose estimated "true" concentration is T. Accordingly, the corresponding total bias ($\hat{\delta}$) is calculated by:

$$\hat{\delta} = \bar{\bar{x}} - T \quad (6)$$

Assuming that the "true" concentration T of the VS has a negligible uncertainty, the variance of the bias can be easily calculated from the ANOVA results [30]:

$$S_{\hat{\delta}}^2 = S^2(\bar{\bar{x}}) = \frac{S_{IP}^2 - ((n-1)/n)S_r^2}{p} \quad (7)$$

We can apply the Student's t-test for assessing the significance of bias:

$$t_{\hat{\delta}} = \frac{\hat{\delta}}{S_{\hat{\delta}}} = \frac{\bar{\bar{x}} - T}{S_{\hat{\delta}}} \quad (8)$$

Table 2
Acceptable recovery percentages according to the concentration level of analyte.

Analyte (%)	Analyte fraction	Concentration unit	Recovery range (%)
100	1	100%	98–102
10	10^{-1}	10%	98–102
1	10^{-2}	1%	97–103
0.1	10^{-3}	0.1%	95–105
0.01	10^{-4}	100 ppm	90–107
0.001	10^{-5}	10 ppm	80–110
0.0001	10^{-6}	1 ppm	80–110
0.00001	10^{-7}	100 ppb	80–110
0.000001	10^{-8}	10 ppb	60–115
0.0000001	10^{-9}	1 ppb	40–120

If $t_{\hat{\delta}}$ is less than the critical tabulated value for $np-1$ degrees of freedom at the chosen significance level, then no statistical differences have been observed between the overall mean and the "true" value.

But, as indicated above, the recovery term can also be used for checking trueness and has a more intuitive meaning. The total recovery for VS is defined as the ratio between the observed estimation of the VS concentration, $\bar{\bar{x}}$ and the "true" value T, expressed as percentage or as fraction:

$$R = \frac{\bar{\bar{x}}}{T} \tag{9}$$

From Eqs. (6) and (8) we get

$$R = 1 + \frac{\hat{\delta}}{T} \tag{10}$$

Thus, if the bias is negative (positive), the recovery will be lesser (higher) than 100. The variance of the recovery is easily obtained and related to those of bias:

$$S_R^2 = \frac{S^2(\bar{\bar{x}})}{T^2} = \frac{S_{\hat{\delta}}^2}{T^2} \tag{11}$$

Once recovery is computed, we can check it for suitability by comparison with the published acceptable recovery range as a function of the analyte concentration [28] as it is depicted in Table 2.

This later procedure is suitable for trueness and intermediate precision assessment coming from a nested ANOVA performed with VS, where trueness and intermediate precision are checked independently. However, it is also possible to assess the accuracy in a global way, with the use of the called accuracy profiles mentioned above.

3. Assessment of global accuracy according to accuracy profiles

The concept of accuracy profile was first introduced in the papers of Hubert et al. [31] and Boulanger et al. [32]. The Societé Française des Ciences et Techniques Pharmaceutiques (SFCTP) have used accuracy profiles to assess the accuracy in method validation according to the concept of acceptability limit [33–35]. Rozet et al., in his excellent paper [36] illustrate very suitably the concepts of trueness, precision, accuracy and accuracy profiles in the sense of SFCTP.

When applying an analytical method, the analyst expects that the difference between the measurement result X and the unknown 'true' value T of the tested sample be less than a predefined acceptance limit λ:

$$-\lambda < X - T < \lambda \Leftrightarrow |X - T| < \lambda \tag{12}$$

The acceptance limit is not arbitrary but depends on the goals of the analytical procedure. It is the outcome of a discussion between the analyst and the client or end-user [12] or is linked to the requirements usually admitted by the practice, for instance, 1–2% on bulk materials, 5% on drug products or pharmaceutical formulations, 15% for biological samples, and so forth [36]. A valid analytical method should provide with results X that accomplish Eq. (11) very likely. This can be formally expressed as:

$$P(|X - T| < \lambda) \geq \beta \tag{13}$$

where β is the probability of having measurements inside the acceptance limits, e.g. 90%. If the true bias (δ) and the true precision (σ) are known, and assuming a normal distribution, we can write $X - T = \delta \pm z_\beta \sigma$, z_β being the β-quantile of the standard normal variate. Eq. (11) can now be expressed as

$$-\lambda < \delta \pm z_\beta \sigma < \lambda \tag{14}$$

But this is a utopia because the true performance parameters are unknown. It is advisable to use their estimations; $\hat{\delta}$ for the bias and $S = \sqrt{S_{IP}^2 + S_{\hat{\delta}}^2}$ for the total precision standard deviation [8].

$$S = \sqrt{S_{IP}^2 + \frac{S_{IP}^2 - ((n-1)/n)S_r^2}{p}} \tag{15}$$

Thus the β-Expectation Tolerance Interval (βETI) can be constructed according to

$$\begin{aligned} \hat{\delta} + kS &> -\lambda \\ \hat{\delta} - kS &< \lambda \end{aligned} \tag{16}$$

where k is the called coverage factor, that can be assimilated to the β-quantile of the standard normal variate, when a Gaussian distribution is assumed. For each VS concentration level, the upper and lower tolerance interval limits, expressed as percentage, are given by

$$\begin{aligned} \text{upper limit}: &\quad \frac{100(\hat{\delta} + kS)}{T} \\ \text{lower limit}: &\quad \frac{100(\hat{\delta} - kS)}{T} \end{aligned} \tag{17}$$

We can construct one interval for each VS. Then, the upper limits of the intervals are connected by straightlines and the lower limits too, leading to two segmented lines. The intersections between these two segmented lines with the acceptance limit straight lines $y = \lambda$ and $y = -\lambda$ (expressed in %) leads to upper and lower quantification limits [5,10].

The excellent e.noval software (Arlenda, Liège, Belgium) can be used to obtain the accuracy profiles and the validation results of the analytical method [37]. Some authors, aside from the variance of intermediate precision and of estimated bias, consider also the robustness variance as a contribution to the total precision variance [5,10,38]. Robustness tests can be considered as intra-laboratory simulations of inter-laboratory studies, if the introduced deliberated variations in the method parameters are suitably selected. Accordingly, the robustness uncertainty can be easily obtained as a relative deviation [16], RSD_{rob}, and can be joined to the budget of the total precision variance RSD.

$$RSD = \sqrt{RSD_{IP}^2 + RSD_{\hat{\delta}}^2 + RSD_{rob}^2} \tag{18}$$

However, the robustness contribution to the total precision variance is generally disregarded by the authors and Eq. (15) is used instead of (18).

4. Conclusion

Intra laboratory testing of method accuracy, when VS are available can be easily performed from the results of one-way ANOVA designed experiments. Then bias and intermediate precision can be

tested independently or in a global manner according to the concept of accuracy profile and β-Expectation Tolerance Interval once the acceptance limit has been selected.

Acknowledgement

Authors are grateful to the Junta de Andalucía (Spain) for grant excellence research project P06-FQM-02029.

References

[1] ICH Harmonised Tripartite Guideline, Validation of Analytical Procedures: Text and Methodology, ICH Working Group, November 2005. http://www.ich.org/LOB/media/MEDIA417.pdf.
[2] Guidance for Industry: Analytical Procedures and Methods Validation, U.S Department of Health and Human Services, Food and Drug Administration, Center for Drug Evaluation and Research (CDER), Center for Biologics Evaluation and Research (CBER), August 2000. http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/ucm122858.pdf.
[3] EURACHEM Guide, The Fitness for Purpose of Analytical Methods, A Laboratory guide to Method Validation and related topics, EURACHEM Working Group, December 1998. http://eurachem.org/guides/valid.pdf.
[4] IUPAC, Pure Appl. Chem. 74 (2002) 835–855.
[5] A.G. González, M.A. Herrador, Trends Anal. Chem. 26 (2007) 227–238.
[6] T.P.J. Linsinger, Trends Anal. Chem. 27 (2008) 916–923.
[7] A.G. González, M.A. Herrador, A.G. Asuero, Talanta 48 (1999) 729–736.
[8] International Organization for Standardization (ISO), ISO 5725-1. Accuracy (trueness and precision) of measurement method and results. Part 1. General principles and definitions, ISO, Geneva, Switzerland, 1994.
[9] A. Menditto, M. Patriarca, B. Magnusson, Accred. Qual. Assur. 12 (2007) 45–47.
[10] A.G. González, M.A. Herrador, Talanta 70 (2006) 896–901.
[11] Ph. Hubert, J.J. Nguyen-Huu, B. Boulanger, E. Chapuzet, P. Chiap, N. Cohen, P.A. Compagnon, W. Dewé, M. Feinberg, M. Lallier, M. Laurentie, N. Mercier, G. Muzard, C. Nivet, L. Valat, E. Rozet, J. Pharm. Biomed. Anal. 45 (2007) 70–81.
[12] V.R. Meyer, J. Chromatogr. A 1158 (2007) 15–24.
[13] M. Feinberg, J. Chromatogr. A 1158 (2007) 174–183.
[14] M. Feinberg, B. Boulanger, W. Dewé, P. Hubert, Anal. Bioanal. Chem. 380 (2004) 502–514.
[15] M. Feinberg, M. Laurentie, Accred. Qual. Assur. 11 (2006) 3–9.
[16] A.G. González, M.A. Herrador, A.G. Asuero, Talanta 65 (2005) 1022–1030.
[17] F.T. Peters, H.H. Maurer, Accred. Qual. Assur. 7 (2002) 441–449.
[18] F.T. Peters, O.H. Drummer, F. Musshoff, For. Sci. Int. 165 (2007) 216–224.
[19] 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, The Netherlands, 1997.
[20] International Organization for Standardization (ISO), ISO-5752-2, Accuracy (Trueness and Precision) of Measurement Methods and Results. Part 2. Basic Method for the determination of Repeatability and Reproducibility of a standard Measurement Method, ISO, Geneva, Switzerland, 1994.
[21] T. Luping, B. Schouenborg, Methodology of Inter-comparison Tests and Statistical Analysis of Test results. Nordtest Project No. 1483-99, SP Swedish National Testing and Research Institute, SP report 2000:35, Boras, Sweden, 2000.
[22] M. Thompson, The Amazing Horwitz Function, AMC Technical Brief No. 17, Royal Society of Chemistry, Cambridge, UK, July, 2004.
[23] R. Wood, Trends Anal. Chem. 18 (1999) 624–632.
[24] W. Horwitz, Anal. Chem. 54 (1982) 67A–76A.
[25] I. Chen, R.R. Eitenmiller, J. Food Chem. Toxicol. 72 (2007) C243–C247.
[26] W. Horwitz, Validation: An invisible component of measurement, AOAC Int., Gaithersburg, MD, 2003. http://www.aoac.org/dietsupp6/Dietary-Supplement-web-site/HorwitzValid.pdf.
[27] E. Pritchard, Quality in the Analytical Chemistry Laboratory. ACOL Series, Wiley, Chichester, West Sussex, UK, 1995.
[28] L. Huber (Ed.), Validation and Qualification in Analytical Laboratories, Interpharm Press, East Englewood, CO, USA, 1998.
[29] AOAC Peer Verified Methods Program. Manual on policies and procedures, AOAC Int., 1998. http://www.aoac.org/vmeth/PVM.pdf.
[30] International Organization for Standardization (ISO), ISO/DTS 21748, Guide to the Use of Repeatability, Reproducibility and Trueness Estimates in Measurement Uncertainty Estimation, ISO, Geneva, Switzerland, 2003.
[31] Ph. Hubert, P. Chiap, J. Crommen, B. Boulanger, E. Chapuzet, N. Mercier, S. Bervoas-Martin, P. Chevalier, D. Grandjean, P. Lagorce, M. Lallier, M.C. Laparra, M. Laurentie, J.C. Nivet, Anal. Chim. Acta 391 (1999) 135–148.
[32] B. Boulanger, P. Chiap, W. Dewé, J. Crommen, Ph. Hubert, J. Pharm. Biomed. Anal. 32 (2003) 753–765.
[33] B. Boulanger, W. Dewé, Ph. Hubert, Objectives of pre-study validation and decision rules, in: AAPS Conference. APQ Open Forum, Indianapolis, 2000.
[34] Ph. Hubert, J.J. Nguyen-Huu, B. Boulanger, E. Chapuzet, P. Chiap, N. Cohen, P.A. Compagnon, W. Dewé, M. Feinberg, M. Lallier, M. Laurentie, N. Mercier, G. Muzard, C. Nivet, L. Valat, J. Pharm. Biomed. Anal. 36 (2004) 579–656.
[35] M. Feinberg, B. Boulanger, W. Dewé, Ph. Hubert, Anal. Bioanal. Chem. 380 (2004) 502–514.
[36] E. Rozet, C. Hubert, A. Ceccato, W. Dewé, E. Ziemons, F. Moonen, K. Michail, R. Wintersteiger, B. Streel, B. Boulanger, Ph. Hubert, J. Chromatogr. A 1158 (2007) 126–137.
[37] Arlenda, Laboratory Solutions, Liège, Belgium. http://www.arlenda.com/e-noval.html.
[38] V.J. Barwick, L.R. Ellison, VAM Project 3.2.1, Development and Harmonisation of Measurement Uncertainty Principles. Part D. Protocol for uncertainty Evaluation from Validation Data, Report No: LGC/VAM/1998/088, January, 2000.