Accred Qual Assur (2007) 12:425–434
DOI 10.1007/s00769-007-0278-1

DISCUSSION FORUM

Repeatability: some aspects concerning the evaluation of the measurement uncertainty

Matthias Rösslein · Sergio Rezzonico · Roman Hedinger · Marco Wolf

Received: 25 July 2006 / Accepted: 27 March 2007 / Published online: 9 June 2007

Abstract Various publications stress the importance of the repeatability (i.e. precision) of the calculation of the measurement of uncertainty. We reveal by detailing an example from production control in the pharmaceutical industry that the effect of other influence quantities should not be neglected, because their magnitude is even larger than the contribution of repeatability. We review the role of repeatability within the calculation of measurement uncertainty for several common validation and day-to-day measurement scenarios. They show that measurement models need to consider the measurement sequences of the various scenarios. Otherwise the size and effect of the repeatability might be overestimated. At the end Monte Carlo simulations were used to investigate the determination of the repeatability under certain restrictions. The simulation uncovered a significant bias toward the common formula for calculating the standard deviation when it is based on a duplicated measurement of a sample.

Keywords Measurement uncertainty · Repeatability · Standard deviation · Production control

M. Rösslein (✉) · S. Rezzonico · R. Hedinger
EMPA, Lerchenfeldstrasse 5, 9014 St Gallen, Switzerland
e-mail: matthias.roesslein@empa.ch

M. Wolf
ETH Zürich, Institute for Scientific Computing,
Universitätstrasse 6, 8092 Zürich, Switzerland

Introduction

The most recent version of the ISO/IEC/EN 17025 [1] standard demands that ''Testing laboratories shall have and shall apply procedures for estimating uncertainty of measurement... (Clause: 5.4.6.2)''. Furthermore, numerous local accreditation bodies require more stringent implementation of those clauses. Therefore, many analytical laboratories in industry have started their own efforts to implement laboratory instructions for estimating the measurement uncertainty. These calculations are based on the description of their various measurement procedures [2].

ISO 17025 (2005) [1] suggests in note 3 of clause 5.4.6.3 that the reader use ISO 5725 [3] and the guide to the expression of uncertainty in measurement (GUM) [4] for further information. In addition, ILAC [5] recommends that analytical laboratories use the Eurachem/CITAC guide ''Quantifying Uncertainty in Analytical Measurement 2nd (QUAM)'' [6]. These guidelines place emphasis to a greater or lesser extent on the repeatability, i.e. closeness of the agreement between the results of repeat measurements of the same measurand carried out under the same conditions of measurement [7]. This overall performance figure is often determined either during the method validation or from the duplicated measurements of the same samples, which are analysed in the daily routine. Various authors of different publications point out that according to their observations repeatability is often the most important component of the calculation of measurement uncertainty in analytical chemistry [2, 8–10]. Hence, it is quite tempting to use only repeatability as a value for measurement uncertainty with the argument that all other components do not have a significant impact on the calculation of the combined standard uncertainty [2]. This attempt is quite attractive, especially for the industry, which has to

fulfil new requirements, but wants to achieve them at minimal costs. For this reason we think it is important to take a closer look at some aspects of repeatability. Within the current article we zero in on the detailed measurement model to check the effect of which influence quantities are covered by the measured repeatability of the analytical procedure and which are not. For this purpose we elaborate on the different details of an example from production control in the pharmaceutical industry. In addition, we look at basic statistical principles, which are often neglected for the calculation of the standard deviation and which might lead to a considerable underestimation of its value and with it of measurement uncertainty.

Example: production control in the pharmaceutical industry

Nowadays, pharmaceutical production often involves more than ten synthesis steps to obtaining an active agent. Hence, it is important to control the production efficiency of each single step within close margins. In addition, the content of the by-products needs tight monitoring to exclude any side effects that might be caused by these impurities. These and other safety risks have led to elaborate risk management within a strictly regulated environment that places high demands on the validation of methods and their implementation in the analytical laboratories [11]. We elaborate within this section on a typical example of an analytical method, which determines the content of the main product and of all relevant by-products obtained from a synthesis step.

Due to the highly competitive nature of pharmaceutical production it is not possible for us to publish any actual analytical procedures or any performance data, such as repeatabilities etc. As an alternative we have utilised an artificial analytical method that is based on numerous similar methods. These have been examined in detail during a major project to develop a software product, which permits the bench chemist to easily calculate measurement uncertainty within a reasonable amount of time [12]—a much-needed tool, because a large company might carry out a few thousand analytical procedures that fall within the scope of ISO 17025.

A syntheses step should result in at least 98% of the content of the key product and five by-products where the content should not exceed 2% for one by-product, 0.5% for two by-products and 0.2% for two by-products. High performance liquid chromatography (HPLC) is the technique of choice to quantify such compound bodies in pharmaceutical industry. The content of the key product is determined by weighing 100 mg of the product and 100 mg of its reference substance and then diluting it with 100 ml of the mobile phase.

The content of each of the 0.5% by-products is quantified with the same measuring solution of the product that was used to determine the content of the key product. The corresponding reference stock solution is produced from 50 mg of the reference standard, which was dissolved in 100 ml of the mobile phase. The measuring solution of the reference was then produced by diluting 1 ml of the stock solution with 100 ml of the mobile phase. For the other by-product the concentration level of the measuring solution of the reference was adjusted by increasing or decreasing the amount of weighed reference substance. All the weight measurements are rounded up or down to 0.1 mg.

Figure 1 shows the flow chart of the analytical procedure. The left side summarises the operations with the samples, whereas the right side describes the operations with the reference substance. The additional steps for diluting the reference solution, which are shown with dashed lines, are only needed to quantify the by-products. According to the given information in the specification we derived the cause and effect diagram of the analytical procedure using the equation used to calculate the measurement result and the rules developed by Ellison et al. (Fig. 2) [6, 13].

Before we continue with the discussion of repeatability issues, we want to have a closer look at the current mea-

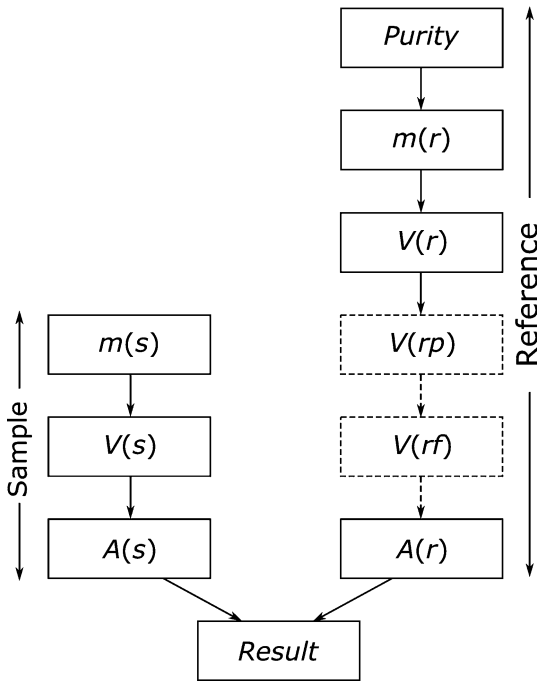

Fig. 1 Operating sequence of the analytical procedure to determine the content of the product and by-products of a synthesis step. *Boxes* with *dashed lines* represent the dilution steps that are needed for the measurements of the content of the by-product

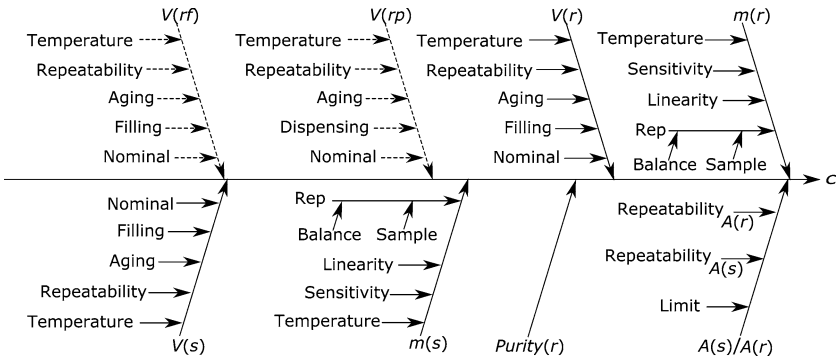

Fig. 2 Cause and effect diagram according to the information of the specification. No influence quantities have yet been rearranged. *Dashed lines* represent parts of the diagram that are only relevant for by-product content determination

surement model displayed as a cause and effect diagram. The individual measurement operations, which are represented in the equation used to calculate the measurement result by different variables, for instance volume, mass, HPLC measurement and purity, follow all the same measurement basics summarised in Fig. 3, i.e. comparison of the sample value with the reference one. The reference values for the measurement operations in our example, with the exception of the HPLC measurements, were not determined in the same laboratory. Therefore, those individual models reflect temporal, local and operational differences with additional influence quantities. For example, the filling of the flask considers personal bias by the operators and possible differences in the formation of the meniscus due to a different solvent in comparison to the water, which was used to calibrate the flask.

The HPLC measurements of the reference and sample signal were made within the same series and on the same instrument. This leads to a different measurement model, where possible differences between the sample and reference measurement and all changes in the reference signal over time are relevant [12], but no significant systematic effects that are equal for the sample and reference measurements. Furthermore, in this example both reference and sample peaks are resolved to base line, which means that there is no difference between the sample and reference signal [12].

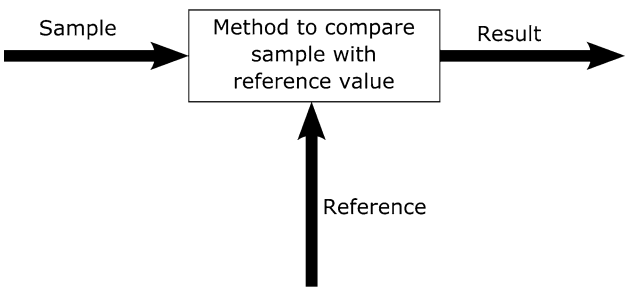

Fig. 3 General measurement principle: method for comparing the sample measurement with the reference measurement

A feasible measurement scenario for the HPLC is shown in Fig. 4. After performing the calibration at $t = 0$ h a calibration control sample is measured at a given time interval of $t = x$ h. If the measured value of the calibration control sample is within a given limit of the value found during the calibration, then the sample measurements are continued, otherwise a new calibration is made before carrying on with the sample measurements.

A typical measurement sequence is displayed in Fig. 5, where first a reference sample is measured followed by duplicate measurements of the samples (samples 1, 2 etc.). The two measurement solutions used for the duplicate measurements on each sample have been prepared inde-

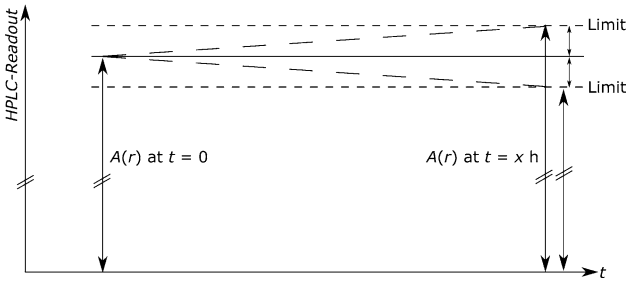

Fig. 4 Calibration control sample: the calibration ($A(r)$ at $t = 0$) remains valid as long as any control measurements ($A(r)$ at $t = x$ h) stay within the given limits

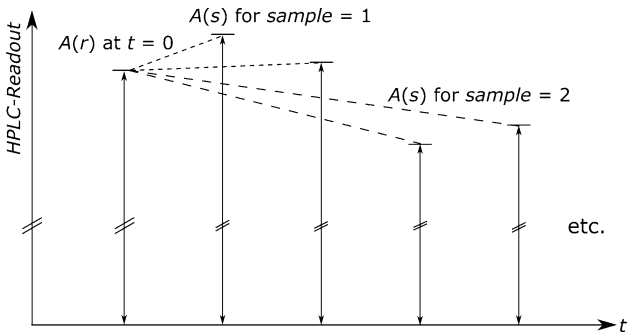

Fig. 5 Measurement sequence for the product and by-product content determination of a synthesis step

pendently. Often the laboratories use repeatability, which was determined from the variation in the duplicate measurements of the same sample, for the calculation of measurement uncertainty [2, 6]. This practice has direct implications for the measurement model. We can only combine those repeatability components of the individual variables in the measurement equation that can vary between duplicate measurements of the same sample. The combination of the influence quantities ''repeatability'' results in a new branch ''repeatability'' representing the overall variation in the results from measurements made of the same sample. Figure 6 shows the modified cause and effect diagram for such a measurement protocol. The repeatability components of all measurement steps related to the sample are combined to the new main branch ''repeatability'' in the diagram. Due to the reasons mentioned above all the other repeatability components, which are indicated by a dashed box in the figure, remain with their individual variables and stay attached to the corresponding main branches. We are of the opinion that this is an important refinement of the general procedure of combining all repeatability components in the diagram, and has been described in the Eurachem/CITAC guide [6]. Further details will be discussed in the next section.

Table 1 summarises the combined standard uncertainty for the main product and for the five by-product content determinations assuming 0.35% overall repeatability and 0.3% standard deviation for the HPLC signal of the measurement solution of the reference. The purity of the main product can be determined with an expanded uncertainty of 1.3% (95% confidence level). This means that the purity of the main product in a given batch has to be at least 99.3% to pass the requirement of better than 98% to take into account measurement uncertainty. The expanded uncertainty of the by-product determinations spans from 1.7 to 2.1% (95% confidence level) depending on the weighed amount of the reference substance. The expanded uncertainty with a level of 95% has been obtained directly from the Monte Carlo distribution

Table 1 Combined standard uncertainty for the product and by-products of a synthesis step

	Product	By-product		
		1%	0.5%	0.2%
u(c)[a] (%)	0.63	0.87	0.90	1.10

[a] Combined standard uncertainty

counting from both sides to determine the value of the appropriate confidence level [12, 14]. If the content of the main component is calculated by summing the total by-product content and subtracting it from 100% purity, then its measurement uncertainty has approximately the same magnitude as that with the direct content determination. This value of measurement uncertainty is the combination of the corresponding measurement uncertainty of all by-product determinations.

The different contributions of the main branches in the cause and effect diagram to the overall measurement uncertainty are shown in Fig. 7. For all investigated cases the overall repeatability is not the major contribution. It is the overall performance of the measurements on the HPLC. One major component is the repeatability of the measurement of the HPLC signal of the reference solution and the other is the size of the boundaries, which are set on the control sample before a recalibration is required. It is important to notice that this variation is not part of the overall repeatability, because the measurement of the reference solution was performed only once (see Fig. 5). Its size has to be determined independently. The pipetting step to dilute the reference stock solution, which is needed for the determination of the by-products, is a relevant contribution, because its small volume of 1 ml requires the use of a micropipette. The weighing of the sample and reference is of minor importance, except for the content determination of the lowermost by-product. Here only 20 mg of the reference standard is weighed to obtain a measurement solution of the reference within the proper target concentration and the read-off value is rounded up or down to 0.1 mg.

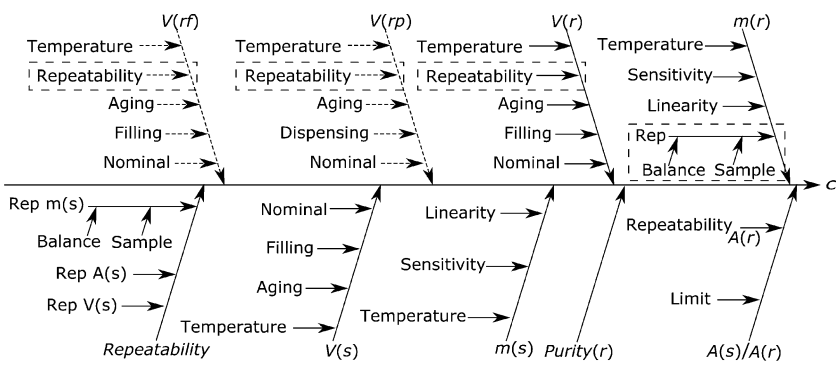

Fig. 6 Cause and effect diagram after rearranging some influence quantities (repeatabilities) to the new main branch repeatability, which represents the variation in the analytical procedure under repeatability conditions. *Dashed boxes* highlight influence quantities (repeatabilities), which cannot be moved due to the measurement sequence

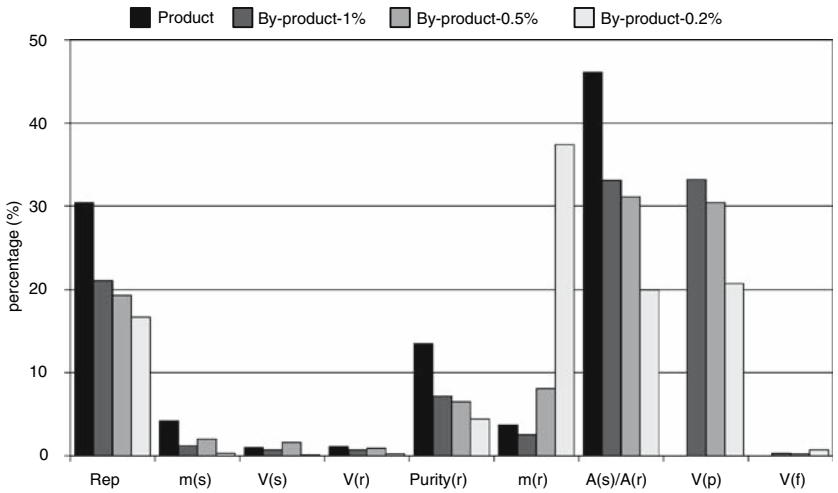

Fig. 7 Contribution of the main effects (main branches in the cause and effect diagram) to the overall measurement uncertainty for the product and by-products (1, 0.5 and 0.2%)

Aspects of repeatability important for the evaluation of measurement uncertainty

If we look at the example discussed in the previous section, then one would expect to be able to develop other measurement scenarios to reduce the value of measurement uncertainty.

Scenario 1

In a first scenario the value for the repeatability of the measurement results obtained using the measurement procedure that has been determined during an in-house validation study is used in the assessment of uncertainty of routine measurements. One condition for using the validation data in this way is the equivalence of the performance of the day-to-day measurements to that of the measurements made during the validation study. There are different approaches to achieving this objective [15].

The most important performance parameter is the repeatability of the measurement result. Therefore, one often observes in the industry the approach to multiplying the value for the repeatability of the measurement result determined during the validation study by a factor of 2.8 ($\approx 1.97 \cdot \sqrt{2}$). For the daily routine those duplicated measurements are then repeated, where the difference between the two measured values is larger than this previously set limit. Of course, the measurement models of the routine analysis and of the validation study have to be the same for this comparison.

If the two measurement models are not equivalent then the following pitfall might unintentionally be disregarded. When performing the validation study all repeatability contributions of the reference and sample branch (see Fig. 1) are combined in the overall repeatability of the measurement result. However, for daily routine work only the sample branch is repeated when making the duplicated sample measurement, which means that only the repeatability components of the sample branch contribute to the repeatability of the measurement result. In this case fewer repeatability contributions are combined in the repeatability of the measurement and therefore its value is most likely smaller than that obtained during the validation study. In other words, if we use now the value of the repeatability of the measurement results determined during the validation study as a benchmark, then for routine work the combined repeatability contributions of the sample branch, which add up to the repeatability of the measurement, are allowed to vary more than during the validation. This leads of course to a larger overall measurement uncertainty compared with one calculated using the data from the validation study. There is an additional pitfall for this type of benchmarking, which will be discussed in the last part of this article.

The other, more indirect approaches to ensuring a similar performance during routine measurements and the validation study cannot overcome the demand for equivalent measurement models. They often have additional drawbacks like the system suitability test [16, 17], which is performed at the beginning of a measurement series and therefore cannot detect any step by step or sudden deterioration of the actual sample measurements.

Scenario 2

In a second scenario the repeatability of the measurement result for an interlaboratory study of an analytical procedure that has been determined according to ISO 5725 [3] is utilised. Here similar arguments are relevant to those in the previous scenario where the values for the evaluation of measurement uncertainty were determined during an in-house validation study. If a laboratory wants to use data

from an interlaboratory test to derive its measurement uncertainty, then the study report has to point out in detail which influence quantities vary between the measurements made by the individual participants. There needs to be a detailed measurement model from all the participants reflecting the measurement sequence and the way to calculate their results. Only under these terms can a laboratory use the within-laboratory standard deviation (s_r) to control the proper implementation of the validated analytical procedure. In order to comply with that requirement the laboratory has to conduct the day-to-day analysis using the same measurement model, which also includes the measurement sequence and the method of calculating the results. In addition, the same combination of sample and reference measurements has to be used. If fewer reference measurements are made then the benchmark repeatability should be reduced, otherwise a larger standard deviation could be accepted for fewer repeatability influence quantities combined to form the new main branch representing the repeatability of the measurement result (see Figs. 5, 6). Furthermore, under this condition the value of the measurement uncertainty of the day-to-day analysis could be larger than that derived from the interlaboratory test, which was conducted at the end of the validation study. It results in a larger overall measurement uncertainty. We observe here similar features to those described in the previous scenario.

Scenario 3

The last scenario is shown in Fig. 8. Two or more reference measurements, which have been determined over the whole measurement series, are combined to a mean value $(\bar{A}(r))$ before the results of the duplicated sample measurements are calculated. The measurement model here is different compared with the previous scenarios. Single influence quantities, which describe the repeatability of individual steps in the reference branch (see Fig. 1), can now be moved to the new main branch ''repeatability'' in the cause and effect diagram. If all measurement solutions are taken from the same solution V(r) or V(rf) respectively, then only the repeatability of the HPLC measurement (A(r)) can be repositioned. On the other hand, if each measurement solution of the reference is prepared from scratch using reference material from at least two manufacturers, who certify their reference substance independently, then all influence quantities of the reference branch, which cover the short-term variation, have to be relocated at the main branch ''repeatability'' of the cause and effect diagram. The current scenario is often representative of a thorough validation study, which is performed in the specialist research laboratories of a larger company. After that, the analytical procedure is transferred to field laboratories, which conduct the day-to-day measurements. The research centres set up a number of limits that allow monitoring the proper implementation of the transferred analytical procedure to the field laboratories. One of these limits is set for the overall repeatability of the measurement results and has been taken from the extensive validation study. The field laboratories try to carry out all their measurements at minimal cost and for that purpose they reduce the total number of measurements as much as possible by cutting down the effort for the measurement of the reference solution. These marginal conditions lead to a measurement sequence that is most likely not comparable with the thorough validation study, but with that described for the example detailed in the previous section. Furthermore, the measurement models for the day-to-day measurements in the field laboratories are not the same as for the elaborate validation study implemented at the research centres. These conditions can lead to an underestimation of the overall measurement uncertainty for the day-to-day measurements, when the variation in the duplicated measurements of the same sample is monitored with the limit set by the results of the validation study. Here we follow the same arguments described for the previous scenarios.

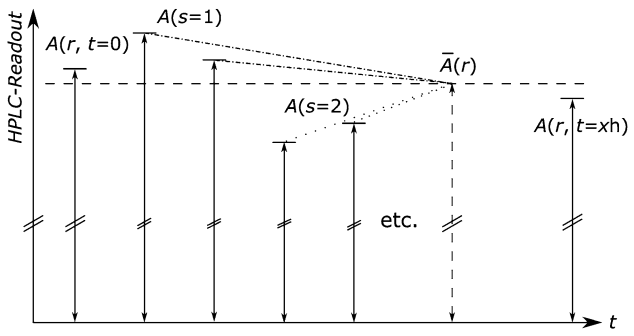

Fig. 8 Measurement sequence and quantification schema for duplicated measurements of the same sample, which refer to more than one measurement of the reference during the measurement series

Duplicated measurements

Finally, we look at the possible direct effects caused by the determination of the standard deviation from the duplicated measurements of the same samples on the measurement uncertainty. It is crucial for the proper quantification of the repeatability that the system has enough time to vary over nearly its whole repeatability range between the individual measurements, i.e. the interval between the two measurements of a duplicate has to be at least as long as the period of all the major frequencies that compose the repeatability of the measurement result [18, 19]. If this requirement is neglected, then the duplicated measurements might be

auto-correlated. Such a possible pitfall is the difference between randomly arranged measurement solutions and a sequence that was set up in chronological order. To illustrate this situation we ran a computer simulation, which is outlined schematically in Fig. 9. In the simulation we generate random numbers that follow a normal distribution with a standard deviation of one. A first random number, which represents the first measurement, is chosen. Its location is given by the probability density function (PDF) of a normal distribution. This normal distribution corresponds to all major frequencies adding up to the repeatability of the measurement result. Then a second random number is selected, which has to be located within the limits given by plus and minus half the range (r) from the position of the first random number. If the second random number is outside the range, then new random numbers are generated until one lies inside the given limits. In this way we simulate two measurements, which take place within a short amount of time and therefore the whole system has no chance to vary over the full extent of the normal distribution. Then the limiting range was broadened step by step until the range covered six times the standard deviation, which corresponds for a normal distribution to 99.7% of its area. At this level the simulation corresponds to a Monte Carlo method (MCM) of a normal distribution. The results of the whole simulation process are summarised in Fig. 10. First we noticed that the calculated mean standard deviation ($\bar{s}$) from the duplicated measurements of the same sample ($n = 2$) grows gradually, but levels off to a value of 0.75. It never reaches the value of 1, which was used as input for the standard deviation while generating the normally distributed random numbers. In other words the 1 million random numbers that form a normal distribution result in a 25% smaller average standard deviation than the input value, only because they were paired, and then the mean standard deviation is calculated from the resulting 500,000 doublets. We repeated the same simulations for $n = 3, 4, 6, 10$ and observed a steady approach towards the

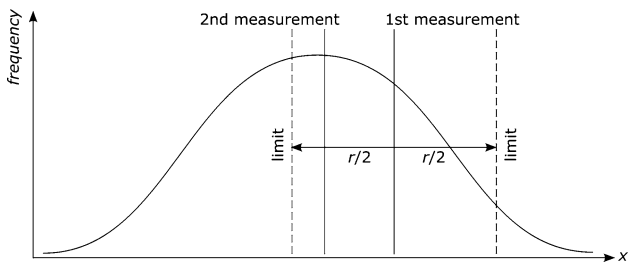

Fig. 9 Schema depicting the determination of the standard deviation for a duplicated measurement whereby the second measurement has to be located within given limits. By this means we simulate a system with a shorter period between the two measurements than the main variation of the investigated framework

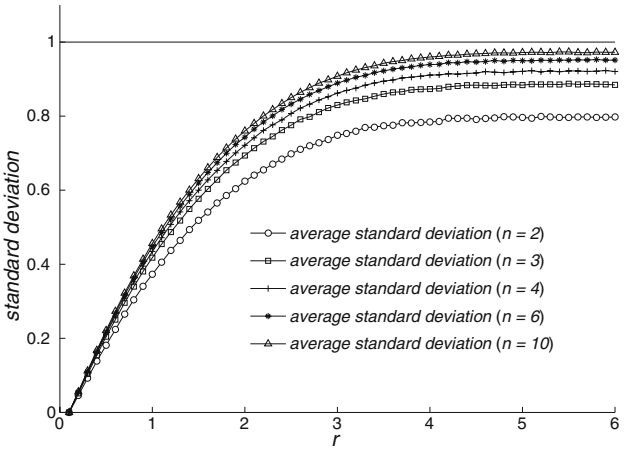

Fig. 10 Results of the Monte Carlo simulation to determine the mean standard deviation for $n = 2, 3, 4, 6, 10$ (n = number of measurements) with a restricted possibility to vary. The mean standard deviation for duplicated measurement exhibits the largest bias at $r = 6$, whereas it approaches step by step that for an increasing number of measurements

expected value of 1. Since this behaviour was not expected, we started looking for an obvious explanation. First, we checked the implementation of the simulation in MatLab [20], but did not find any error. Then we started looking at some textbooks on statistics and we found in a number of them [21–23] the explanation we were looking for. The common formula (Eq. 1), which is generally used to calculate the standard deviation, delivers a biassed estimate of the standard deviation for small numbers of measurements.

$$s_x = \sqrt{\frac{1}{n-1}\sum_{I=1}^{n}(x_I - \bar{x})^2} \tag{1}$$

with

s_x biassed estimate of standard deviation of a single measurement
n total number of measurements
x_I I-th element in the number of measurements with the value x
$\bar{x}$ arithmetic mean of the n measurements with values x

The formula (Eq. 2) [22, 23], which delivers an unbiased estimate of the standard deviation, is of special importance to small numbers of measurements.

$$\hat{s}_x = \sqrt{\frac{n-1}{2}}\frac{\Gamma\left(\frac{n-1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)}s_x \tag{2}$$

with

$\hat{s}_x$ unbiased estimate of the standard deviation
n total number of measurements

Table 2 Correction factor to adjust the biassed standard deviation of a given sample size (n)

Sample size (n)	Correction factor
2	1.253314
3	1.128379
4	1.085402
5	1.063846
10	1.028109
15	1.018002

$\Gamma(n)$ gamma function
s_x biassed estimate of the standard deviation of a single measurement

Table 2 summarises the correction factors that adjust the values of the biased estimate of the standard deviation to those of the unbiased one (see Eq. 2). It is strictly only correct for normally distributed observations. (see Appendix). We were surprised to find that this fact was not described anywhere in the relevant ISO standards, especially in ISO 5725 [3].

We have to recall that often bench chemists use the mean standard deviation of duplicated sample measurements as repeatability of the measurement result (e.g. Eurachem/CITAC QUAM2002 example 4 [6]) and this value is biassed according to the results presented above. Therefore, the effect of the biassed estimate of the standard deviation on the calculation of measurement uncertainty should be thoroughly investigated in the near future. As elaborated previously, the repeatability of an analytical procedure is an important component, but not the only one, of the calculation of measurement uncertainty. If the value of the repeatability is based on a biassed estimate of the standard deviation due to duplicated measurements of a sample, and if it builds the largest contribution to the overall measurement uncertainty, then the combined standard uncertainty is up to 25% too small, leading to an expanded uncertainty ($k = 2$) that is also 25% too small. This fact has a direct impact on compliance testing and has considerable consequences when the results are compared with the given limits of other results. All effects and resulting consequences of the biassed estimate might not be that obvious in all cases. For example, the repeatability of an analytical procedure was determined during an in-house validation study of ten repeated measurements. The validation study was designed in such a way that it followed the same measurement model as that for the day-to-day measurements of duplicated samples. Hereby, the design obeys the demands that have been elaborated in the previous sections of this article. But still the measurement uncertainty of the day-to-day measurements might be considerably larger than that calculated from the validation data, because the biased estimate of the standard deviation allows the duplicated sample measurements to vary by up to 22% (see Table 2) more than one would expect from a limit set up according to the results of the validation study.

Finally, we look at the primary objectives of the simulation after its unexpected aspects have been explained. All the mean standard deviations (see Fig. 10) show a steady increase with the increasing size of the range, which limits the possible location of the second ($n = 2$) measurement or each subsequent measurement ($n = 3, 4, 6, 10$). The limiting value of the mean standard deviation is approximated to a size of the range r equals four, which corresponds to a variation between minus and plus two standard deviations. This means that the interval between the two measurements of each duplicated sample has to be long enough to allow a full variation over the whole distribution of the results. Otherwise, the determined standard deviation is too small and its value gets larger or even smaller if the timing of the measurement sequence is prolonged or shortened respectively. This leads to widened boundaries for the acceptance of results from individual samples. Currently, there are no propositions or models that allow that effect to be incorporated into the calculation of measurement uncertainty.

Conclusion

A number of publications stress the fact that the repeatability of the measurement result obtained using an analytical procedure is a major contribution to its measurement uncertainty. A typical example representing production control in the pharmaceutical industry reveals that other contributions are at least as important as the repeatability of the measurement result and that their omission leads to a significant underestimation of the overall measurement uncertainty. Furthermore, the standard deviation of an analytical procedure covers only those influence quantities that had the possibility to vary during the period of the individual measurements of the same sample. For this reason the model of the measurement, which can be elaborated as a cause and effect diagram, has to consider the measurement sequence during the validation study and the day-to-day measurements. Otherwise, the size of the measurement uncertainty might be underestimated considerably because performance parameters of a different nature are compared. Those characteristics are illustrated describing a number of common validation and day-to-day measurement scenarios.

Monte Carlo simulations, which were started to determine the behaviour of the standard deviation from the duplicated measurement of the same samples under certainty restrictions, displayed unexpected results. They

highlighted the fact that the mean standard deviation determined from just two measurements of the same samples is strongly biassed. This bias might lead to an underestimation of the expanded uncertainty ($k = 2$) of up to 25%. In addition, we were able to show with the help of the simulation that the period between the individual measurements of the same sample needs to be long enough so that the influence quantities, which comprise the overall repeatability of the analytical procedure, have enough time to vary over the full range.

Acknowledgement We would like to thank Alex Williams to help us improve the revised version considerably.

Appendix

Estimate of the standard deviation

The following deduction was added, because it is relatively difficult to find it in most textbooks about statistics, and it is mainly based on [22].

Looking for the density $h(s)$ of the random variable $S = \sqrt{\frac{1}{n-1}\sum_{I=1}^{n}(X_I - \bar{X})}$.

The random variable $Y = \frac{(n-1)S^2}{\sigma^2}$ follows a Chi-squared distribution with n –1 degrees of freedom for a normally distributed random variable. According to [21] (Chap. IV, Sect. 5.8.1) the random variable Y holds the density $g(y) = 0$ for $y < 0$ and

$$g(y) = \frac{1}{2^{(n-1)/2}\Gamma\left(\frac{n-1}{2}\right)}\exp(-\frac{y}{2})y^{(n-3)/2} \quad \text{for } y \geq 0.$$

At first the density $w(s^2)$ of the random variable S^2 is determined. In this process s^2 is the variable and not s. One obtains from $g(y)$ with substitution

$$y = \frac{n-1}{\sigma^2}s^2; \mathrm{d}y = \frac{n-1}{\sigma^2}\mathrm{d}s^2$$

$$w(s^2) = \frac{1}{2^{\frac{n-1}{2}}\Gamma\left(\frac{n-1}{2}\right)}\exp\left(-\frac{(n-1)s^2}{2\sigma^2}\right)\left(\frac{n-1}{\sigma^2}s^2\right)^{(n-3)/2}\frac{n-1}{\sigma^2}$$
$$= \frac{(n-1)^{\frac{n-1}{2}}}{2^{(n-1)/2}\Gamma\left(\frac{n-1}{2}\right)\sigma^{n-1}}\exp\left(-\frac{(n-1)s^2}{2\sigma^2}\right)(s^2)^{(n-3)/2}$$

One obtains from it using the substitution $s = \sqrt{s^2}; s^2 = (s)^2; \mathrm{d}s = 2s\mathrm{d}s$ the density $h(s)$ of the random variable S in the form of

$$h(s) = \frac{(n-1)^{(n-1)/2}}{2^{(n-1)/2}\Gamma\left(\frac{n-1}{2}\right)\sigma^{n-1}}\exp\left(-\frac{(n-1)s^2}{2\sigma^2}\right)s^{n-3}2s$$
$$= \frac{(n-1)^{(n-1)/2}}{2^{(n-3)/2}\Gamma\left(\frac{n-1}{2}\right)\sigma^{n-1}}\exp\left(-\frac{(n-1)s^2}{2\sigma^2}\right)s^{n-2} \quad \text{for } s > 0$$

The expected value of the random variable is obtained as

$$E(S) = \int_0^\infty s\,h(s)\mathrm{d}s$$
$$= \frac{(n-1)^{\frac{n-1}{2}}}{2^{\frac{n-3}{2}}\Gamma\left(\frac{n-1}{2}\right)\sigma^{n-1}}\int_0^\infty s^{n-1}\exp\left(-\frac{(n-1)s^2}{2\sigma^2}\right)\mathrm{d}s$$

This integral is transformed using the substitution $\frac{(n-1)s^2}{2\sigma^2} = z; s = \sqrt{\frac{2z}{n-1}}\sigma; \frac{(n-1)s}{\sigma^2}\mathrm{d}s = \mathrm{d}z; \mathrm{d}s = \frac{\sigma}{\sqrt{n-1}\sqrt{2z}}\mathrm{d}z$ to

$$E(S) = \frac{(n-1)^{(n-1)/2}}{2^{(n-3)/2}\Gamma\left(\frac{n-1}{2}\right)\sigma^{n-1}}\int_0^\infty\left(\frac{2\sigma^2 z}{n-1}\right)^{\frac{n-1}{2}}$$
$$\times\exp(-z)\frac{\sigma}{\sqrt{n-1}\sqrt{2z}}\mathrm{d}z$$
$$= \frac{\sqrt{2}\sigma}{\sqrt{n-1}\Gamma\left(\frac{n-1}{2}\right)}\underbrace{\int_0^\infty z^{\frac{n}{2}-1}\exp(-z)\mathrm{d}z}_{=\Gamma\left(\frac{n}{2}\right)}$$
$$= \frac{\sqrt{2}\Gamma\left(\frac{n}{2}\right)}{\sqrt{n-1}\Gamma\left(\frac{n-1}{2}\right)}\sigma < \sigma$$

It follows based on the Stirling equation that

$$\lim_{n\to\infty}\frac{\Gamma\left(\frac{n}{2}\right)}{\sqrt{n-1}\Gamma\left(\frac{n-1}{2}\right)} = \frac{1}{\sqrt{2}}$$

It is for this reason that

$$\lim_{n\to\infty}E(s) = \sigma$$

The estimating function S is for σ asymptotically unbiased.

Its variance is

$$\mathrm{Var}(S) = E(S^2) - [E(S)]^2$$

$$\sigma^2 - \frac{2\Gamma^2\left(\frac{n}{2}\right)}{(n-1)\Gamma^2\left(\frac{n-1}{2}\right)}\sigma^2 = \left(1 - \frac{2\Gamma^2\left(\frac{n}{2}\right)}{(n-1)\Gamma^2\left(\frac{n-1}{2}\right)}\right)\sigma^2$$

The estimating function

$$\hat{S} = \sqrt{\frac{n-1}{2}}\frac{\Gamma\left(\frac{n-1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)}S = \frac{1}{\sqrt{2}}\frac{\Gamma\left(\frac{n-1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)}\sqrt{\sum_{i=1}^{n}(X_I - \bar{X})^2}$$

is unbiased with normally distributed random variables and with any kind of sample size n

$$E(\hat{S}) = \sigma;$$

$$\mathrm{Var}(\hat{S}) = \frac{n-1}{2}\frac{\Gamma^2\left(\frac{n-1}{2}\right)}{\Gamma^2\left(\frac{n}{2}\right)}\mathrm{Var}(S) = \left(\frac{n-1}{2}\frac{\Gamma^2\left(\frac{n-1}{2}\right)}{\Gamma^2\left(\frac{n}{2}\right)} - 1\right)\sigma^2$$

with

$$\lim_{n\to\infty}\frac{n-1}{2}\frac{\Gamma^2\left(\frac{n-1}{2}\right)}{\Gamma^2\left(\frac{n}{2}\right)} = 1.$$

References

1. ISO (2005) ISO/IEC/EN 17025: general requirements for the competence of testing and calibration laboratories. ISO, Geneva
2. Populaire S, Campos Giménez E (2006) A simplified approach to the estimation of analytical measurement uncertainty. Accred Qual Assur 10:485–493
3. ISO (1994) ISO 5725: accuracy (trueness and precision) of measurement methods and results. ISO, Geneva
4. ISO (1993) Guide to the expression of uncertainty in measurement. ISO, Geneva
5. ILAC (2002) ILAC-G17: introducing the concept of uncertainty of measurement in testing in application of the standard ISO/IEC 17025. ILAC, Silverwater
6. Ellison SLR, Rösslein M, Williams A (eds) (2000) Eurachem/CITAC guide: quantifying uncertainty in analytical measurement, 2nd edn. Eurachem, Teddington
7. JCGM Working Group on International Vocabulary of Basic and General Terms in Metrology (1993) International vocabulary of basic and general terms in metrology (VIM). ISO, Geneva
8. Armishaw P (2003) Estimating measurement uncertainty in an afternoon. A case study in the practical application of measurement uncertainty. Accred Qual Assur 8:218–224
9. ISO (2004) ISO/TS 21748: guidance for the use of repeatability, reproducibility and trueness estimates in measurement uncertainty estimation. ISO, Geneva
10. Horwitz W, Kamps LR, Boyer KW (1980) Quality assurance in the analysis of foods and trace constituents. J Assoc Off Anal Chem 63:1344–1354
11. US Department of Health and Human Services (2000) Guidance for industry—analytical procedures and methods validation. US Department of Health and Human Services, Food and Drug Administration, Center for Drug Evaluation and Research (CDER), Center for Biologics Evaluation and Research (CBER), Rockville
12. Uncertainty Manager Development Core Team (2006) Uncertainty manager—evaluation of the measurement uncertainty in analytical chemistry. St. Gallen, Switzerland
13. Ellison SLR, Barwick VJ (1998) Estimating measurement uncertainty: reconciliation using a cause and effect approach. Accred Qual Assur 3:101–105
14. JCGM Working Group on the Expression of Uncertainty in Measurement (2004) Draft-guide to the expression of uncertainty in measurement. Supplement 1: numerical methods for the propagation of distributions
15. Kromidas S (2000) Handbuch Validierung in der Analytik. Wiley/VCH, Weinheim
16. Dong M, Paul R, Gershanov L (2001) Getting the peaks perfect: System suitability for HPLC. Todays Chem Work 10:38–40, 42
17. US Department of Health and Human Services (1996) Guidance for industry—Q2B validation of analytical procedures: methodology. US Department of Health and Human Services, Food and Drug Administration, Center for Drug Evaluation and Research (CDER), Center for Biologics Evaluation and Research (CBER), Rockville
18. Wegscheider W (2007) Universitätslehrgang 'Qualitätssicherung im chemischen Labor' Modul D: Messunsicherheit und Prozessanalytik. Montanuniversität Leoben, Leoben
19. Moser J, Wegscheider W, Meisel T, Fellner N (2003) An uncertainty budget for trace analysis by isotope-dilution ICP-MS with proper consideration of correlation. Anal Bioanal Chem 377(1):97–110
20. MatLab Development Core Team (1994–2006) The MathWorks, Natick
21. Sachs L (1992) Angewandte statistik, 7th edn. Springer, Heidelberg
22. Bolch BW (1968) More on unbiased estimation of the standard deviation. Am Stat 22:27
23. Bosch K (1993) Statistiktaschenbuch, 2nd edn. Oldenbourg, Munich