Available online at www.sciencedirect.com
SCIENCE DIRECT®
Journal of Pharmaceutical and Biomedical Analysis 37 (2005) 859–870
JOURNAL OF PHARMACEUTICAL AND BIOMEDICAL ANALYSIS
www.elsevier.com/locate/jpba

Validation in pharmaceutical analysis☆ Part II: central importance of precision to establish acceptance criteria and for verifying and improving the quality of analytical data

Joachim Ermer[a,*], Hans-Joachim Ploss[b]

[a] *Quality Operations, Global Analytical Development, Aventis, Frankfurt, Germany*
[b] *Quality Operations Drug Products, Aventis, Frankfurt, Germany*

Received 20 May 2004; accepted 26 June 2004
Available online 14 August 2004

Abstract

Validation of analytical procedures is a vital aspect not just for regulatory purposes, but also for their efficient and reliable long-term application. In order to address the performance of the analytical procedure adequately, the analyst is responsible to identify the relevant parameters, to design the experimental validation studies accordingly and to define appropriate acceptance criteria.

Establishing an acceptable analytical variability for the given application is of central importance as many other acceptance criteria can be derived from such a precision. Acceptable precision ranges for types of control tests and/or analytes can be obtained from validation, but also related activities such as transfer, control charts, or extracted from routine applications such as batch release or stability studies (data mining). Apart from compiling a database for general benchmarking, during such an information-building process, the reliability of the analytical variability of the specific procedure is more and more increased. This is important as a reliable target variability facilitates to detect or investigate atypical or out-of specification behaviour of analytical data in a routine application, thus improving the data quality and reliability.

According to the life-cycle concept of validation, measures should be taken to maintain and control the validated status of analytical procedures during long-term routine application, such as monitoring relevant performance parameters (system suitability tests), control charts, etc.

If the analytical system is demonstrated to be stable, i.e. under statistical control, a major variability contribution in LC originating from the standard preparation and analysis can be reduced. A concept of quantification by pre-determined calibration parameters instead of the classical approach of simultaneous calibration is described.

Keywords: Acceptance criteria; Analytical variability; Precision; Pre-determined calibration; Quality assurance; Validation

1. Introduction

Accuracy and reliability of the analytical results is crucial for ensuring quality, safety, and efficacy of pharmaceuticals. Consequently, analytical validation has been in the focus of regulatory requirements for a long time [2–5]. However, a sensible validation is also essential from a business perspective because analytical data are the basis of many decisions such as batch release, establishment/verification of shelf life, etc.

ICH guidelines should be regarded as basis and philosophy of analytical validation, not as a checklist. "*It is the responsibility of the applicant to choose the validation procedure and protocol most suitable for their product*" [2b]. Suitability is strongly connected with the requirements and the design of the given analytical procedure, which obviously

☆ Part I: see reference [1].

* Corresponding author. Tel.: +49 69 305 84890; fax: +49 69 305 20522.
E-mail address: joachim.ermer@aventis.com (J. Ermer).

0731-7085/$ – see front matter © 2004 Elsevier B.V. All rights reserved.
doi:10.1016/j.jpba.2004.06.018

varies and must, therefore, be reflected in the analytical validation. This includes the identification of the performance parameters relevant for the given procedure, the definition of appropriate acceptance criteria, and the appropriate design of the validation studies. In order to achieve this, the analyst must be aware of the fundamental meaning of these performance parameters, calculations, and tests and their relationship to his specific application. A lack of knowledge or (perhaps) a wrong understanding of "efficiency" will lead to validation results that address the real performance of the analytical procedure only partly or insufficiently. In the best case, it is a waste of resources because the results are meaningless.

The suitability of an analytical procedure is primarily determined by the requirements to the given test item, and secondarily by its design (that is normally more flexible). Usually, the (minimum) requirements are defined by the acceptance limits of the specification (specification limits). For some applications, the requirements are explicitly defined in ICH guidelines. For example, the reporting level for unknown degradants in drug products (DP) is set to 0.1 and 0.05% for a maximum daily intake of less and more than 1 g active, respectively [2d]. Consequently, the corresponding test procedure must be able to quantify impurities at this concentration with an appropriate level of precision and accuracy.

Many other performance parameters are linked with the analytical variability. Therefore, once an acceptable precision is defined, it can serve as an orientation for other acceptance criteria. As far as possible, normalised (percentage) parameters should be defined as validation acceptance limits because they can be compared across methods, and therefore more easily drawn from previous experience. Statistical significance tests should very cautiously be (directly) applied as acceptance criteria because they can only test for a statistical, not a practical significance. On one hand, due to the small number of data normally used in pharmaceutical analysis, large confidence intervals may obscure not acceptable differences. On the other hand, in case of a larger number of data (as a consequence of tightening confidence intervals) or by sometimes abnormally small variability in one of the analytical series (that is, however, of no risk for routine application), differences are identified as significant which are of no practical relevance [6]. In addition, when comparing independent methods for the proof of accuracy, different specificities can be expected which add a systematic bias, thus increasing the risk of the aforementioned danger.

It is also important to understand validation as a life-cycle approach [7]. The user of any method has to guarantee that it will stay consistently in a validated status. This does not necessarily mean that additional work needs to be performed. During the application of the analytical procedures, a lot of data providing performance information are generated, e.g. system suitability tests, repeated measurements, etc., but often left unused. In order to make rational and efficient use of these data, they must be processed, eventually enabling to achieve better understanding and control of the analytical procedure or process.

2. Analytical variability and validation parameters

2.1. Precision

2.1.1. Minimum requirements imposed by the specification limits

Generally, acceptance limits of the specification have to enclose both the analytical and the manufacturing variability [2e]. The former was considered as a confidence interval [1,8]. This concept was refined to address the analytical variability by prediction intervals [9] because future applications are the more appropriate objective. Rearranging this equation, the maximum permitted analytical variability for assay can be calculated from the acceptance limits of the specification Eq. (1).

$$\mathrm{RSD}_{\max}(\%) = \frac{|(\mathrm{BL} - \mathrm{SL})| \times \sqrt{n_{\mathrm{assay}}}}{t(P, \mathrm{d}f)} \quad (1)$$

where SL: acceptance limits of the specification for active ingredient (% label claim); BL: basic limits, 100%—maximum variation of the manufacturing process or decrease in stability (in percentage). The BL closest to the respective SL must be used; n_{assay}: number of repeated, independent determinations in routine analyses (if the mean is compared to the acceptance limits. If each individual determination is defined as the reportable result, $n = 1$ has to be used); $t(P,\mathrm{d}f)$: Student t-factor for the defined level of statistical confidence (usually 95%) and the degrees of freedom in the respective precision study.

In contrast to the method capability index where the normal distribution is used to describe the range required for the analytical variability, Eq. (1) can take variable number of determinations directly into account as well as the reliability of the experimental standard deviation (by means of the Student t-factor). Of course, the precision acceptance limit thus obtained will be the minimum requirement. If a tighter control is aimed at or if a lower variability is expected for the given type of method, the acceptance limits should be adjusted accordingly (see Section 2.1.4).

2.1.2. Precision levels and variability of the standard deviation

Regarding an analytical procedure, each of the steps will contribute with its variability, usually summarized as the precision levels system (or instrument) precision, repeatability, intermediate precision, and reproducibility. Each of the levels includes the lower ones, as well as numerous individual steps. The analyst must be strictly aware of these levels and the importance of their correct reporting. If, for example, a repeated injection of the same test solution were wrongly reported as repeatability, the whole sample preparation would be ignored, that may be an important contribution to this precision level. Reproducibility is defined as between-laboratory

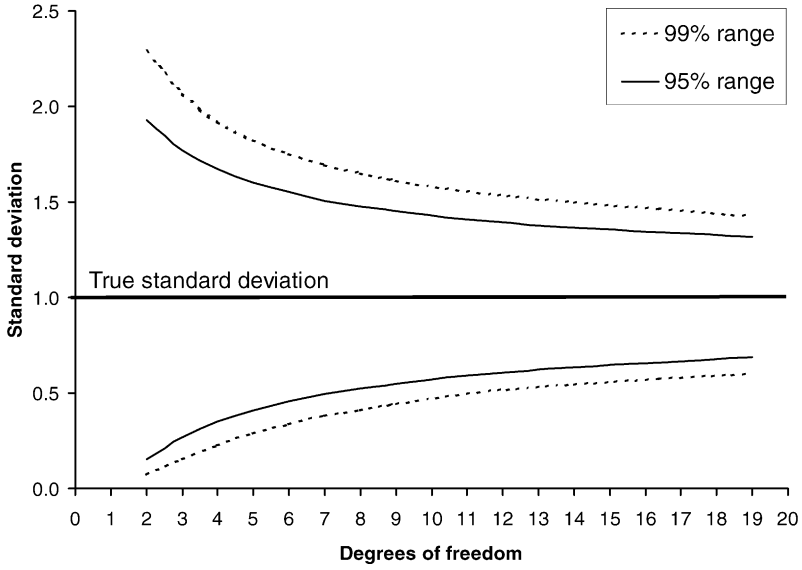

Fig. 1. Ranges of standard deviations calculated from 50.000 simulated normally distributed data sets in dependence on their sample size. The mean and the standard deviation used for the simulation were 100 and 1, respectively. The lower and upper limits are shown between the indicated percentage of the calculated standard deviations was found.

precision [2a], but in the long-term perspective, both intermediate precision and reproducibility approach each other, at least in the same company. Therefore, in the following both (sub)levels are discussed in combination. Reproducibility from collaborative trials can be expected to include additional contributions due to a probably larger difference of knowledge, experience, equipment, etc. among the participating laboratories.

The analyst should also take into consideration that experimental standard deviations show a large variability. In Fig. 1, the lower and upper limits of the distribution ranges of simulated standard deviations for various numbers of determinations are shown. The smaller the number of data, the higher is the variability. For small numbers of data, the standard deviation distribution is skewed towards higher values because the lower side is limited by zero. Standard deviations calculated from six values (five degrees of freedom) were found up to the 1.6-fold of the true value, taking 95% of all results into account. This is important to realize when acceptance criteria for experimental standard deviations are to be defined as here the upper limit of their distribution is relevant.

2.1.3. Acceptable precision ranges for LC assay – system or injection precision

In the European Pharmacopoeia, for chromatographic assay of drug substance (DS), a maximum permitted system precision is defined, in dependence on the upper specification limit (USL) and the number of injections [10]. The difference between the upper specification limit and 100% corresponds to the range available for the analytical variability because the content of a drug substance cannot be larger than 100%. An analytically available range of 2.0%, for example, allows a relative standard deviation of 0.73 and 0.85%, for five and six injections, respectively. The FDA [4] and Canadian guidelines [3] recommend system precisions less than 1.0%. In order to reflect really the instrument (i.e. mainly the injection) precision, the concentration must be sufficiently larger (at least 100-fold) than the quantification limit; otherwise the integration or noise error will have a substantial contribution [11,12].

2.1.4. Acceptable precision ranges for LC assay – repeatability and intermediate precision

2.1.4.1. Experimental. Two hundred and twenty-four individual series from reversed-phase LC assay determinations of 28 different drug substances, formulated in 36 drug products were compiled. The data originate from validation studies, analytical transfers, stability studies (see Section 3.2), and reference standard retests. Repeatabilities of individual data sets were calculated if at least 5 repetitions were performed, the average number of repetitions is 7.2. If two and more series were available, the overall repeatability (Eq. (2)) and intermediate precision/reproducibility (Eq. (4) or Eq. (6)) were calculated [13,14], as well as the ratio between the precision levels. In order to minimize the influence of extreme results, these were omitted for the calculation of averages and the empirical estimation of ranges of the respective subgroup. Only single extreme values were omitted, the percentage of results included is shown in Tables 1–3 in the column "Range" (in brackets). The average relative standard deviations were calculated taking the respective degrees of freedom into account (weighing).

$$\text{Overall repeatability} \quad s_r^2 = \frac{\sum(s_j^2)}{k} \quad s_r = \sqrt{s_r^2} \qquad (2)$$

Table 1
Repeatabilities for the investigated types of drug product

Drug product type (no. in Fig. 2)	No.[a]	Average (%)[b]	Range% (%)[c]	Ratio[d]
Drug substance (no. 1–8)	45 (2)	0.50	0.20–0.96 (96)	1.9
Lyophilisates (no. 9–14)	31 (2)	0.56	0.10–1.00 (94)	1.8
Others (no. 15–19)	11 (1)	0.63	0.22–0.99 (92)	1.6
Solutions and suspensions (no. 20–27)	53 (2)	0.47	0.13–1.18 (96)	2.5
Tablets and capsules (no. 28–44)	64 (4)	0.81	0.16–1.51 (94)	1.9
Combined drug substance, lyophilisates, solutions, and others (no. 1–27)	141 (4)	0.52	0.12–1.18 (97)	2.3

[a] Number of individual repeatabilities used for evaluation (omitted extreme results).
[b] Weighted, pooled standard deviation.
[c] Range of the individual repeatabilities (% of results included).
[d] Between upper range limit and average.

Table 2
Intermediate precisions/reproducibilities for the investigated types of drug product

Drug product type (no. in Fig. 3)	No.[a]	Average (%)[b]	Range% (%)[c]	Ratio[d]
Drug substance (no. 1–7)	15 (1)	1.05	0.35–1.68 (94)	1.6
Lyophilisates (no. 8–12)	7 (2)	0.77	0.38–1.29 (78)	1.7
Solutions and suspensions (no. 16–121)	23 (2)	0.69	0.39–1.05 (92)	1.5
Tablets and capsules (no. 22–29)	18 (0)	1.38	0.42–2.34 (100)	1.7

[a] Number of reproducibilities used for evaluation (omitted extreme results).
[b] Weighted, pooled standard deviation.
[c] Range of the intermediate precision/reproducibilities (% of results included).
[d] Between upper range limit and average.

Table 3
Ratio between reproducibility and overall repeatability

Drug product type (no. in Fig. 4)	No.[a]	Average	Upper limit (%)[b]
Drug substance (no. 1–7)	14 (2)	1.9	3.0 (88)
Lyophilisates (no. 8–12)	8 (1)	1.6	1.9 (89)
Solutions and suspensions (no. 16–21)	23 (2)	1.7	2.6 (92)
Tablets and capsules (no. 22–29)	17 (1)	1.8	2.7 (94)

[a] Number of ratios used for evaluation (omitted extreme results).
[b] Percentage of results included.

$$\text{Inter-group variance} \quad s_g^2 = \frac{\sum\left(\bar{x}_j^2\right) - \frac{\left(\sum \bar{x}_j\right)^2}{k}}{k-1} - \frac{s_r^2}{n} = s_{\bar{x}}^2 - \frac{s_r^2}{n} \quad (\text{if } s_g^2 < 0 \Rightarrow s_g^2 = 0) \tag{3}$$

$$\text{Intermediate precision/reproducibility} \quad s_R^2 = s_r^2 + s_g^2, \quad s_R = \sqrt{s_R^2} \tag{4}$$

where s_j, $\bar{x}_j$: standard deviation and mean of series j; n: number of determinations per series (identical for all); k: number of series; and $s_{\bar{x}}$: standard deviation of the means.

2.1.4.2. Results and discussion. The results show no clear dependency on the analyte, but rather on the type of drug product (see Tables 1–3, Figs. 2–4). However, the target values and especially the distribution ranges discussed should be regarded as orientation for typical applications, it can be expected that some analytes/methods require larger variabilities. The average value for each subgroup can be regarded as a target standard deviation (TSD) [15], i.e. an estimate for the true variability for this group. The limits of the range can serve as orientation for a maximum acceptable variability.

DS and DP, apart from tablets, show similar distributions of repeatabilities, with a range from 0.1 to 1.0–1.2% and an average of 0.52%. This TSD corresponds well to the result of 0.6% from a collaborative trial of the European Pharmacopoeia for the LC assay of cloxacillin [16]. The LC-assay for tablets and capsules is accompanied with a larger variability range from 0.2 to 1.5% and a higher average of 0.81%. This is probably caused by the more complex sample and/or sample preparation. For both groups, the ratio between the upper limit and the average repeatability is about 2, which corresponds very well with the upper 95% confidence limit of a standard deviation (2.1 for $n = 6$ and 1.9 for $n = 7$).

In the range investigated, neither a log–linear correlation of the variability to the concentration fraction (between 0.001 and 0.99) nor to the amount of analyt injected (between 0.8 and 15 μg) is observed.

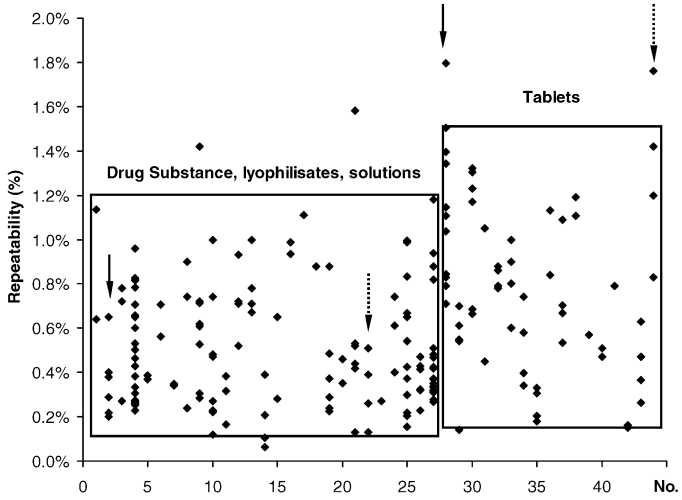

Fig. 2. Repeatabilities for 28 actives in 36 drug products and 8 drug substances. The number on the *x*-axis corresponds to the different analytes per drug product (for details, see Table 1). Arrows indicate the same active in different drug products.

In case of intermediate precision/reproducibility, the averages are between 1.4 and 2.1 times larger than for repeatability, reflecting the additional variability contributions, such as reference standard, time, operator, equipment, laboratory, etc. From the DP with sufficient data, three subgroups for intermediate precision/reproducibility can be defined: DS, tablets/capsules and a combined group of lyophilisates, solutions/suspensions, and others with an upper reproducibility distribution limit of 1.7, 2.3, and 1.1–1.3%, respectively. The ratio between upper limit and average reproducibility is with 1.6, 1.7, and 1.6, respectively, slightly smaller than for the repatabilities. This may be explained by the larger number of determinations, the reproducibility determinations are based on.

For practical purposes, the factors between the precision levels, namely between repeatability and reproducibility, are very important. A classification of these factors would allow to predict the long-term variability of given analytical procedures from repeatability determinations. The calculations are shown in Table 3 and Fig. 4. It must be taken into consideration that the uncertainty of the ratio is larger because it includes the uncertainty of both precision levels.

The smallest possible ratio is 1.0, i.e. no additional variability between the series is observed and both precision levels have the same standard deviation. Experimentally, this can occur even if the true ratio is larger than 1 if one or several experimental repeatabilities are obtained in the upper range

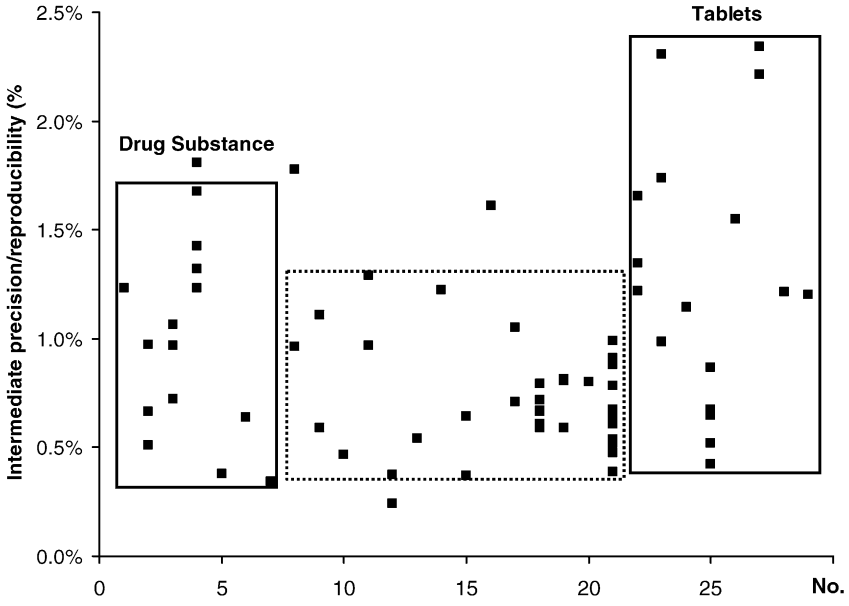

Fig. 3. Intermediate precisions/reproducibilities for 22 actives in 22 drug products and 7 drug substances. The number on the *x*-axis corresponds to the different analytes per drug product (for details, see Table 2).

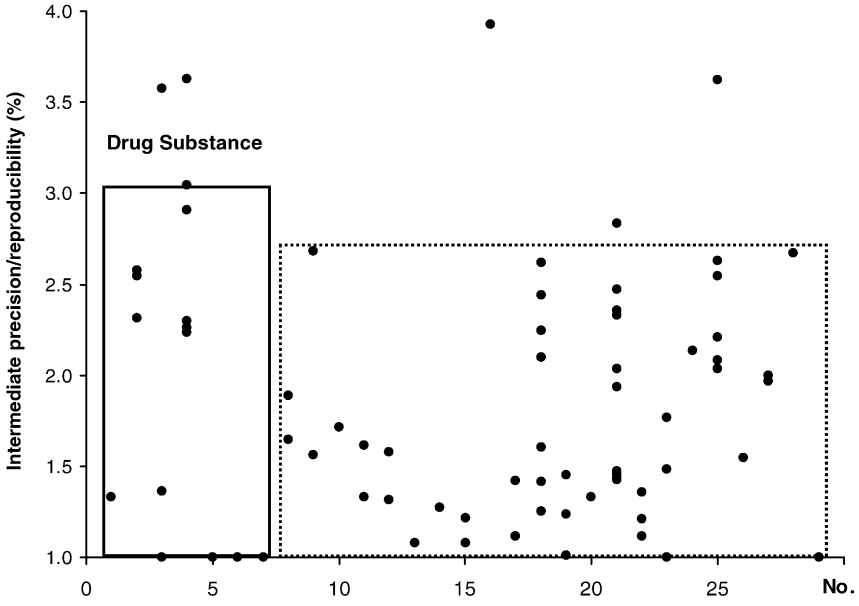

Fig. 4. Ratio between reproducibility and overall repeatability for for 22 actives in 22 drug products and 7 drug substances. The number on the *x*-axis corresponds to the different analytes per drug product (for details, see Table 3).

of the distribution, thus covering the differences between the series.

The upper distribution limit of the factors was not much different for DS and DP, with 3.0 and 2.7, respectively. In between DP, there seems to be no obvious differences, apart from lyophilisates with a more narrow distribution. However, the number of data obtained is not sufficient to draw this conclusion reliably. For DS, the larger upper factor may be explained by the simple sample preparation. As a consequence, the influence of the reference standard to the overall variability is increased, affecting directly the reproducibility.

As a generalisation, the relation between the precision levels can be expressed as multiples of the repeatability target standard deviation:

- Acceptable individual repeatability <2 × TSD.
- Acceptable overall repeatability <1.5 × TSD.
- Acceptable intermediate precision <3–4 × TSD.

These results are in good agreement with the more general estimation of factors between the precision levels of about 1.5 per level [17], i.e. a ratio of 2.2 for repeatability and long-term precision.

2.2. Accuracy

A maximum acceptable difference may be derived from statistical considerations. The *t*-test can be regarded as the description of the relationship between a difference (between two means or to a reference) and a standard deviation. Rearranging the corresponding equations, the maximum permitted difference is given as a function of the (maximum permitted) standard deviation (Eq. (5)). The factors F depend only on the number of determinations and whether the comparison is to a nominal value or another experimental result. Under the usually applied conditions, i.e. six or nine determinations, the factors are approximately one. Therefore, an acceptable precision can be used as an orientation for the difference acceptance limit with respect to means.

Relation between precision and difference:

$$\text{to a nominal value :} \quad D \leq \frac{t(P, df)}{\sqrt{n}} \times s = F \times s$$

$$\text{between means} \quad D \leq t(P, f) \times \sqrt{\frac{2}{n}} \times s = F \times s \qquad (5)$$

If an acceptable difference between individual determinations is of interest (e.g. stability of test solutions, trend analysis, etc.), the concept of variability limits Eq. (6) can be used [13]. R describes the maximum range (or difference between two random values) that can be statistically expected (e.g. with a confidence of 95%). As σ, the TSD for the given application can be used. The precision level determines the application of the variability limit, e.g. with a standard deviation of the injection precision, the maximum difference between two injections of the same solution is obtained, with a repeatability the maximum range of independent sample preparations, etc. Such limits can be defined directly as intermediary acceptance criteria within the analytical procedure as a measure of performance verification, or applied in case of investigations of out-of specification or suspect (out-of trend, out-of expectation) results [18].

Variability limit :

$$R_{\mathrm{I,r,R}} = z \times \sqrt{2} \times \sigma = 1.96 \times \sqrt{2} \times \sigma \approx 2.8 \times \sigma_{\mathrm{I,r,R}} \qquad (6)$$

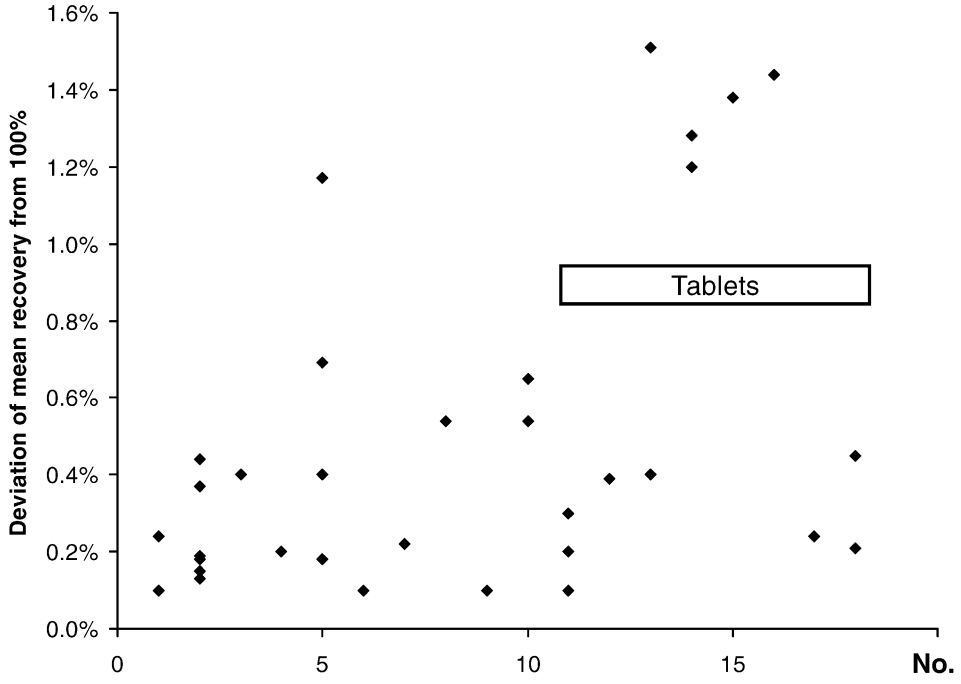

Fig. 5. Difference between mean recoveries of LC-assays and the theoretical value of 100%. The number corresponds to the different analytes per drug product investigated.

where $R_{I,r,R}$: limit according to the corresponding precision level, i.e. injection, repeatability, and reproducibility, respectively.

The theoretically obtained relationship between the maximum permitted precision and the difference between means can also be supported by experimental results. The difference between the mean recovery and the theoretical value of 100% for 36 recovery series for LC assays of 18 drug products are shown in Fig. 5. The usual spiking range of the active into the placebo was 80–120 or 70–130%, the number of determinations ranged from five to nine. If sufficient data were available, the concentration levels are shown separately.

The absolute differences range from 0.1 to 1.5%, with an average bias of 0.53%. Due to the relatively small number of data, further classification according to the type of drug product is not possible, although it seems to be that the deviations from the theoretical value are slightly larger for tablets. The average difference of 0.53% is in perfect agreement with the repeatability obtained for the subgroup of DS and DP requiring less complex sample preparation (Table 1).

2.3. Linearity

The coefficient of correlation is almost uniformly (mis)used, but it is neither a proof of linearity nor a (suitable) quantitative measure [19,20]. In contrast, this parameter requires a linear response function as a prerequisite. In other words, the correlation coefficient requires random scatter around the linear regression line to have any meaning, but even then the numerical values cannot be properly compared because they depend on the slope [21], as well as on the number of determinations and the regression concentration range [12]. Therefore, this parameter is not suitable as a general acceptance criterion for the performance of an analytical procedure.

Usually, in pharmaceutical analysis the intrinsic response or calibration function is known and therefore, the question is rather about a verification, or lack of deviation from linearity. This can be achieved by an analysis of the residuals, i.e. the difference between the experimental and the calculated (from the regression line) response (y-value). A visual evaluation of the pattern of the residuals (versus x or calculated y-values, residual plot) is a very simple and straightforward, but nevertheless powerful tool to detect deviations from the regression model. If the linear, unweighted regression model is correct, the residual plot must show random behaviour in a constant range, without systematic pattern. An acceptable dispersion range for the residuals will correspond to about ± 3 times the expected TSD. For a limited working range, this can be defined with respect to the nominal working concentration (100%). From Table 1, residuals should usually scatter ± 1.5 to 2% around 0. Non-linear behaviour will result in systematic or curved pattern of the residuals, non-constant variances (heteroscedasticity) in a wedge-shaped distribution, with increasing residuals [21].

The standard error of slope (or standard error of regression) measures the deviation of the experimental values from the regression line and thus represents a good performance parameter with respect to the precision of the regression. Expressed in percent (relative standard error of slope), it is comparable to the relative standard deviation obtained in precision studies in the given concentration range. In statistical

textbooks, the parameter is normalised with respect to the mean *x*-value (Eq. (8)). Because the regression range in pharmaceutical analysis is not always centred around the target (working) concentration, the latter can be used instead [22].

Residual standard deviation of regression :

$$s_y = \sqrt{\frac{\sum (y_i - (a + b \times x_i))^2}{n - 2}} \quad (7)$$

Standard error of slope : $$V_{x0} = \frac{s_y}{b \times \bar{x}} \times 100\% \quad (8)$$

The absence of constant systematic errors is a prerequisite for a single point calibration and for the 100% method (area normalisation) for the determination of impurities. The so-called single-point calibration represents, in fact, a two-point calibration line where one point equals zero and the other the standard concentration. This negligible intercept has to be demonstrated experimentally, a regression forced through zero is only justified afterwards.

A negligible intercept can be demonstrated statistically by means of the confidence interval of the intercept, usually at 95% level of significance. If it includes zero, the intercept is statistically not significant. However, a small variability may result in a significant intercept, but without any practical relevance. In contrast, a large variability can obscure a substantial deviation of the intercept from zero. For an absolute evaluation, the intercept can be expressed as a percentage of the analytical signal at the target or a reference concentration, such as 100% working concentration in case of assays. In fact, this approach can be regarded as an extrapolation of the variability at the working concentration to the origin. Therefore, an acceptable precision value can be used as orientation for an acceptance limit. Because large extrapolation will increase the uncertainty of the calculated intercept and may adversely affect its evaluation [12], the minimum range of 80–120% required for an assay is should be extended. Starting from 10 or 20% will avoid extrapolation artefacts as well as maintain the required homogeneity of variances.

3. Sources for long-term precision

Because of the importance of reliable precision data, both for generating a benchmark for orientation (generally, or for relevant classes of methods and/or analytes) and for analytical quality assurance with respect to specific analytical procedures, some examples are given in the following on how to obtain appropriate data. These approaches, when applied regularly, provide at the same time a continuous overview on the on-going performance of the analytical procedure and demonstrate that the validated status is maintained, according to the life-cycle concept of validation [5,7].

Table 4
Calculation of repeatability from duplicate sample preparations of a semi-solid drug product

Batch	Content (mg/g)		
	Preparation 1	Preparation 2	Difference
D261	2.6936	2.7107	−0.0171
D260	2.7224	2.7132	0.0092
D259	2.6896	2.6995	−0.0099
D258	2.7276	2.6958	0.0318
D257	2.7095	2.7258	−0.0163
D256	2.7196	2.7149	0.0047
D255	2.6916	2.7333	−0.0417
D254	2.6986	2.6730	0.0256
D253	2.6760	2.6975	−0.0215
D252	2.6772	2.6721	0.0051
D251	2.6791	2.6728	0.0063
Mean content			2.6997
Standard deviation			0.0146
Relative standard deviation			0.54%

3.1. Batch release

Repeated injections of a standard solution are usually part of the system suitability test of LC assays [10]. The system or injection precision thus obtained can be monitored in a control chart (see Section 3.3) and used to calculate an ongoing average. The control chart will provide also the distribution of the individual system precisions. If the data are recorded with traceability to the LC system, they can be used at the same time to monitor and verify the instrument performance as part of the equipment qualification.

If multiple sample preparations are performed, the repeatability can be calculated from the difference of duplicates or between the minimum and maximum determination (Eq. (9)), an example is shown in Table 4. A prerequisite to calculate a relative standard deviation is a narrow distribution of the content values, which is usually fulfilled in pharmaceutical batch release. Summarizing the differences over all batches analysed results not only in a sufficient number of data to achieve good reliability, but takes also the time aspect into account and provides medium to long-term precision results. The same approach can be used to calculate an average system precision, if the differences between duplicate injections of the same sample solution are used.

Standard deviation from differences :

$$s_d^2 = \frac{\sum \left(x_{i,\max} - x_{i,\min}\right)^2}{2 \times k} \quad s_d = \sqrt{s_d^2} \quad (9)$$

where *k*: number of samples or batches analysed in duplicates.

3.2. Stability studies

In stability studies, the same analytical procedure is applied over a long time. Therefore, these data are an excellent source to provide very reliable, long-term analytical variability. A prerequisite to calculate precision are non-rounded, individual results for each storage interval. If repeated de-

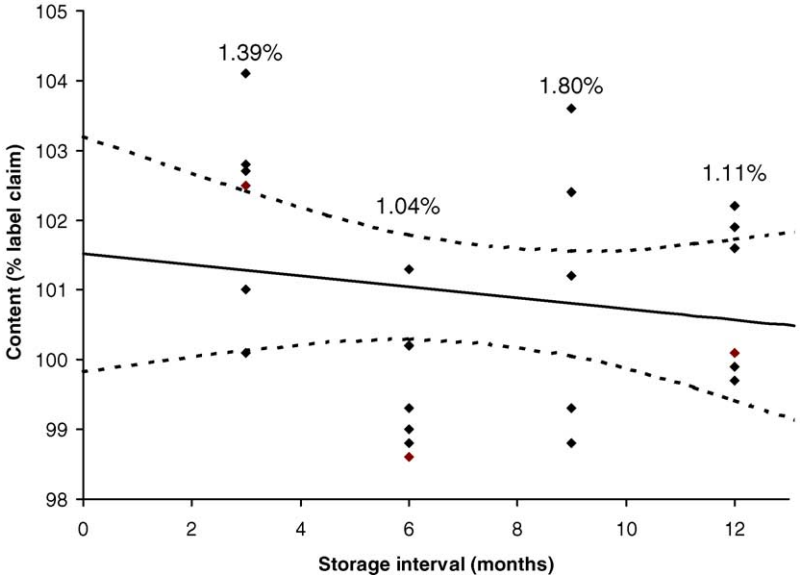

Fig. 6. Example for calculation of precisions from a stability study of film coated tablets. Two presentations per storage interval could be combined. The individual repeatabilities per storage interval and the regression line (solid) with 95% confidence interval (dotted line) are indicated.

terminations are performed for each storage interval, both overall repeatability (according to Eq. (2)) and reproducibility can be calculated, in case of sufficient replicates also individual repeatabilities. In order to increase the number of replicates, several presentations or storage temperatures of the same bulk batch can be combined, if they do not have an influence on the stability and if they were analysed in the same series, using the same reference standard preparations.

Reproducibilities are calculated either using Eq. (4) or – in case of a significant decrease in content – from the residual standard deviation of the linear regression (Eq. (7)) of the individual content determinations (y-values) versus the storage time (x-values). In order to normalize this parameter, it is referred to the content mean.

In the example given in Fig. 6, the overall repeatability was calculated to 1.37%. The confidence interval of the slope includes zero and is not significant. Therefore, the reproducibility can be calculated by an ANOVA (Eq. (4)) resulting in 1.67%. Comparing this result with the residual standard deviation of the regression of 1.62%, both calculation procedures result in identical reproducibilities. However, due to the weighing effect included in the regression and the mean content value, the content decrease should be limited to about 10%.

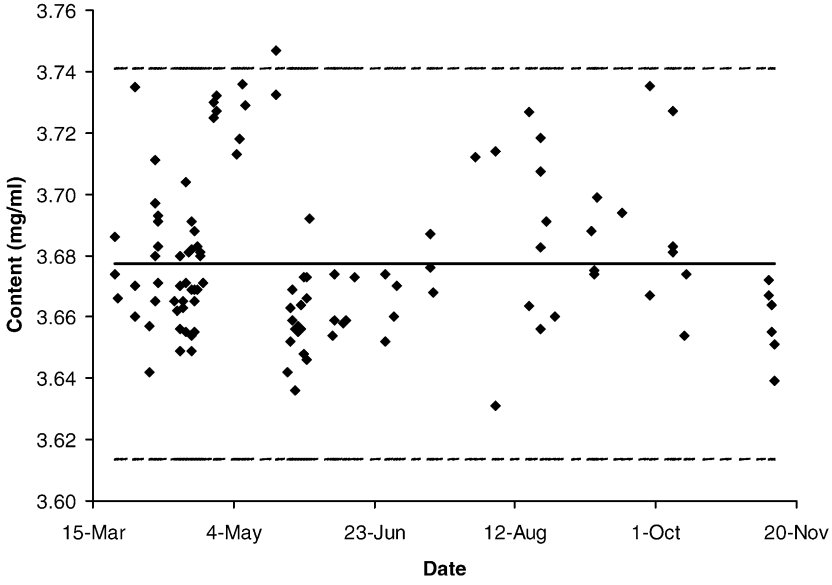

Fig. 7. Control chart for a single determination of an injection solution. The average (solid line) and the $\pm 3\sigma$ control limits (broken line), calculated from the first 20 determinations are indicated. The overall reproducibility was calculated from more than 100 determinations to 0.70%.

3.3. Control charts

Control charts are an excellent tool to inspect the long-term behaviour of the analytical procedure, or individual parts of it. As soon as sufficient information is gathered, they provide information whether the monitored parameter is in (statistical) control or not [23]. At the same time, the data for control charts can be used to calculate on a continuous and successive basis precisions, e.g. a target system precision from a system precision chart or reproducibility from a control sample chart (see Fig. 7).

4. Quantification by pre-determined calibration parameters versus classical approach of simultaneous calibration in LC

Control charts can be used as a tool for identifying root causes for systematic errors on the one hand and – as a consequence – for reduction of variability of an analytical procedure on the other hand. One main root cause for variability in quantitative analysis is the preparation and analysis of an external standard. Depending on its stability, the external standard solution is prepared for each analytical set separately or after certain intervals (e.g. weekly or monthly). The external standard solution is usually analysed on the analytical equipment (e.g. LC-system or spectrometer) daily for each analytical series, i.e. simultaneously with the samples. Based on the results of this determination the parameters for the calibration function are calculated according to the control test for each analytical series separately. These parameters are then used to calculate the content of the analyte in an unknown sample (simultaneous calibration).

While this calibration approach makes the analytical procedure robust for changes in the equipment, it finally leads to an additional contribution to the variance of the result and to an over-adjustment if the system is stable. This additional variance of the external standard can be caused by variation in content of ampouled standards, by weighing or dilution procedures or, in LC, by injection procedures.

An alternative approach calculating the parameters of a calibration function for every analytical series is to take predetermined parameters. The content of an analyte in an unknown sample will then be calculated by a calibration function with these constant parameters. However, this requires verification, whether the pre-determined parameters are still valid for quantification, which is done by analysis of a control sample. The analytical result of the control sample is then checked in a control chart. If it is within pre-determined limits, the calibration is valid and can be used to quantify an unknown sample.

For a high volume drug product (solution for injection), this approach of pre-determined calibration parameters was compared to the traditional approach of simultaneous calibration. The quantitative analysis is performed by means of LC using a reversed-phase column RP18. The nominal content of the analyte in the drug product is 13.2 μg/mL with an upper specification limit of 13.8 μg/mL and a lower specification limit (LSL) of 12.6 μg/mL. According to the traditional approach, the standard solutions are prepared by weighing of a reference standard powder into a flask and by diluting it to volume. Each standard solution is prepared twice and injected twice. Based on the results of the chromatographic analysis, a response factor is calculated. Quantification is done using a single-point calibration and the content of the analyte is determined by the following equation:

$$\text{Content}_{\text{sample}}(\mu\text{g/mL}) = \text{Area}_{\text{sample}}(\text{mV})/\text{Response factor}_{\text{standard}}(\text{mL} \times \text{mV}/\mu\text{g}) \quad (10)$$

In the traditional approach, the response factor was freshly determined for each analytical set. In the new approach, the response factor is kept constant for all analytical series, and verified by the control sample. This constant response factor was calculated as the mean response factor from 48 analytical series in the time period of 11 months to 11544 (mL × mV/μg). In parallel to the determination of the response factor, a control sample was established. The control sample consisted of several hundred units of the drug product taken from one routine manufacturing batch. As the product is a solution for injection, the variation in content between the different units was considered as being close to zero. It therefore describes completely the analytical variation.

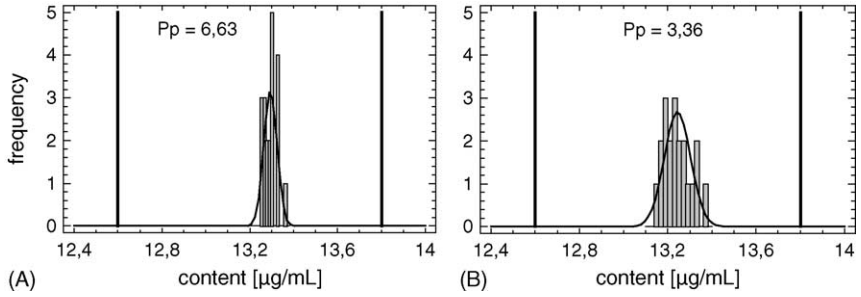

Fig. 8. Process capability for pre-determined (A) and simultaneous (B) calibration. Upper and lower specification limits are indicated by vertical lines. The distribution of the 20 content determinations of the control sample is shown by a histogram and the calculated normal distribution curve [24].

Table 5
Results of 20 control samples by simultaneous and pre-determined calibration

Statistical parameter	Simultaneous calibration	Pre-determined calibration
Count (*n*)	20	20
Mean (μg/mL)	13.244	13.295
Relative standard deviation (%)	0.45	0.23
Minimum (μg/mL)	13.153	13.252
Maximum (μg/mL)	13.361	13.364
Range (%)	1.57	0.84
Test on normal distribution (standardized skewness and kurtosis)	Pass	Pass

For each of the 48 analytical series, one unit of the control sample was analysed in parallel to the batches for release testing. The area values of these determinations were then retrospectively transformed into content by use of the pre-determined response factor. Based on these results the upper and lower control limit for the control chart was determined to 13.19 and 13.48 μg/mL. The design of the new approach implies to use one dedicated LC-system.

After design of the new approach, a test interval of 3 months was conducted. During the test interval 20 analytical series were conducted. Results for the control samples of each series were calculated both by simultaneous and pre-determined calibration.

The results of these determinations are visualised in process capability plots (Fig. 8). The increase in precision for quantification with pre-determined calibration parameters is well described by the improvement of the capability index Pp (Eq. (11)) from 3.36 for the simultaneous calibration to 6.63 for the pre-determined calibration.

$$\mathrm{Pp} = \frac{\mathrm{USL} - \mathrm{LSL}}{6 \times s} \tag{11}$$

The results of the control samples in the 20 analytical series are summarized in Table 5. These results also demonstrate an increase in precision by factor two, expressed in standard deviations. Besides the increase in precision the new approach also saves time. The time consuming preparation of standards is only conducted when the acceptance criteria for the control sample are not met. In such a case, a failure investigation needs to be started. If a root cause cannot be found and/or eliminated (e.g. an instrument shift after repair or equipment qualification), the pre-determined response factor needs to be re-adjusted. In such a case, the simultaneous calibration is performed until the standard deviation of a moving average of sufficient control samples (at least six) meets a certain pre-defined acceptance limit. An acceptable limit should be based on the variance of the simultaneous calibration method, e.g. derived from the 3 months test interval. The new pre-determined response factor is then calculated as the average of these six individual (simultaneous) response factors. Alternatively, it can be obtained from the control chart by means of the ratio between the new control sample average (after the analytical procedure is under control, i.e. with the acceptable variability) and the overall average until the out-of control situation. The old pre-determined response factor is then corrected by this ratio.

5. Conclusions

In order to establish acceptance criteria for validation or transfer of analytical procedures and to evaluate or investigate routine analytical results, it is important to have a reliable expectation of the target variability of the (type of) analytical procedure in question.

For specific methods, the various precisions can be obtained (or refined on an ongoing basis) from batch release results, stability studies, transfers, and control charts, etc. Stability studies especially allow calculating both long-term repeatability and reproducibility.

Collecting such data, target precisions and ranges for groups of drug products or analytes can be estimated. In the present investigation for LC assay, acceptable repeatability, overall repeatability, and intermediate precision/reproducibility can usually be expected to be less than 2, 1.5, and 3 times the repeatability target standard deviation, respectively.

Keeping the analytical procedure under control allows replacing simultaneous calibration by calculation of the analyte content using constant response factors, the so-called pre-determined calibration. Especially for large-volume products, considerable savings are possible, as well as an increase in precision.

Acknowledgements

The presented results originate from the work of numerous of our Aventis colleagues we like to acknowledge here, namely J. Herchen, L. Kuhlmann, S. Lange, T. Moertel, N. Schlicht, M. Schmidt, G. Schumacher-Voigt, I. Thiel-Bielefeld, and S. Weinig.

References

[1] J. Ermer, J. Pharm. Biomed. Anal. 24 (2001) 755–767.
[2] International Conference on the Harmonization of Technical Requirements for the Registration of Pharmaceuticals for Human Use (ICH), http://www.ich.org/UrlGrpServer.jser?@_ID=276&@_TEMPLATE=254.

(a) ICH Q2A: Text on Validation of Analytical Procedures, October 1994.
(b) ICH Q2B: Validation of Analytical Procedures: Methodology, November 1996.
(c) ICH Q3A: Impurities in New Drug Substances, March 1995.
(d) ICH Q3B: Impurities in New Drug Products, November 1996.
(e) ICH Q6A: Specifications: Test Procedures and Acceptance Criteria for New Drug Substances and New Drug Products, Chemical Substances, October 1999.
[3] Acceptable methods, Drug Directorate Guidelines, National Health and Welfare, Health Protection Branch, Health and Welfare Canada, 1992.
[4] CDER Guideline on Validation of Chromatographic Methods, Reviewer Guidance of Chromatographic Methods, US Food and Drug Administration, Centre for Drugs and Biologics, Department of Health and Human Services, 1994.
[5] Draft Guidance Analytical Procedures and Methods Validation. US Food and Drug Administration, Centre for Drugs and Biologics, Department of Health and Human Services, 2000, http://www.fda.gov/cder/guidance/2396dft.htm#III.
[6] USP Draft <1010>, Pharmacopeial Forum 30 (2004) 236–263.
[7] G.C. Hokanson, Pharm. Technol. 18 (1994) 92–100, 118–130.
[8] F.J. Van de Vaart, Pharmeuropa 9 (1997) 139–143.
[9] J. Ermer, H. Wätzig, Regul. Aff. J. (2003) 1–3.
[10] European Pharmacopeia, 4.5 (2004) 2.2.46 Chromatographic Separation Techniques.
[11] U. Schepers, J. Ermer, L. Preu, H. Wätzig, J. Chromatogr. B, in press.
[12] J. Ermer, Performance parameters, calculations and tests, in: J. Ermer, J.H.M. Miller (Eds.), Practical Method Validation in Pharmaceutical Analysis, Wiley VCH, Weinheim, 2004, in press (Chapter 2).
[13] DIN ISO 5725-2: accuracy (trueness and precision) of measurement methods and results; a basic method for the determination of repeatability and reproducibility of a standard measurement method, 1990.
[14] Validation software MVA 2.0 (2001) NOVIA GmbH, Frankfurt, Germany, http://www.novia.de.
[15] A.G.J. Daas, J.H.M. Miller, Pharmeuropa 9 (1997) 148–156.
[16] A.G.J. Daas, J.H.M. Miller, Pharmeuropa 10 (1998) 137–146.
[17] B. Renger, J. Chromatogr. B 745 (2000) 167–176.
[18] FDA guidance for industry: investigating out of specification (OOS) test results for pharmaceutical production (Draft Guidance), 1998.
[19] Analytical Methods Commitee, amc technical brief no. 3, December 2000, http://www.rsc.org/lap/rsccom/amc/amc_techbriefs.htm.
[20] S. Burke, LC-GC Europe Online Supplement statistics and data analysis (2001) 13–18, http://www.lcgceurope.com/lcgceurope/article/articleList.jsp?categoryId=935.
[21] K. Baumann, Process Control Qual. 10 (1997) 75–112.
[22] M. Bloch, Validation during drug product development, in: J. Ermer, J.H.M. Miller (Eds.), Practical Method Validation in Pharmaceutical Analysis, Wiley VCH, Weinheim, 2004, in press (Chapter 5).
[23] D. Shainin, P.D. Shainin, Statistical process control, in: L.M. Juran, F.M. Gryna (Eds.), Juran's Quality Control Handbook, McGraw-Hill, NY, 1988, pp. 24.1–24.39.
[24] Statgraphics Plus Vers. 5.1, Manugistics, Inc., 2115 East Jefferson Street, Rockville, Maryland 20852, U.S.A.