Available online at www.sciencedirect.com
SCIENCE DIRECT®
Journal of Pharmaceutical and Biomedical Analysis 38 (2005) 653–663
JOURNAL OF PHARMACEUTICAL AND BIOMEDICAL ANALYSIS
www.elsevier.com/locate/jpba

Precision from drug stability studies Investigation of reliable repeatability and intermediate precision of HPLC assay procedures

Joachim Ermer[a,*], Christoph Arth[b], Phillippe De Raeve[c], Donald Dill[d], Horst-Dieter Friedel[e], Heidi Höwer-Fritzen[f], Gerd Kleinschmidt[g], Gerhard Köller[h], Heidi Köppel[i], Mathias Kramer[j], Markus Maegerlein[k], Udo Schepers[i], Hermann Wätzig[i]

[a] *Department of Industrial Quality and Compliance, Sanofi-Aventis, Industriepark Höchst, Geb. H831 D-65926, Frankfurt am Main, Germany*
[b] *PDG Analysis, Schwarz BioSciences GmbH, Monheim, Germany*
[c] *Quality Assistance, Donstiennes, Belgium*
[d] *Labor für Analytik und Stabilitätsprüfung, A&M StabTest GmbH, Mainz, Germany*
[e] *Bayer HealthCare AG, Pharma Operations, Quality Assurance, Integrated Quality Management, Leverkusen, Germany*
[f] *Analytical Development and Stability Testing, Spitzner Arzneimittel, Ettlingen, Germany*
[g] *Analytical Sciences Department, Sanofi-Aventis, Frankfurt, Germany*
[h] *Analytical Sciences, Boehringer Ingelheim Pharma GmbH and Co.KG, Biberach a.d. Riss, Germany*
[i] *TU Braunschweig, Institute of Pharmaceutical Chemistry, Braunschweig, Germany*
[j] *QA Stability, Roche, Basel, Switzerland*
[k] *Analytical Development SOLIQS, Abbott, Ludwigshafen, Germany*

Received 9 February 2005; accepted 9 February 2005
Available online 12 April 2005

Abstract

A multi-company investigation is presented to obtain and compare precision results for LC assay procedures. Forty-four drug substances and drug products of various types subjected to 156 stability studies, with 2915 assay values in total, were included. This provides an excellent source of real long-term precision estimates, as the same analytical procedure was applied during the whole stability study, extending from 12 to 60 months. Intermediate precision was calculated either using the residual standard deviation of the regression line or applying an analysis of variances, depending on whether there was a significant degradation of the analyte or not. The results show impressively the large intervals where the individually calculated parameters scatter. Distribution ranges and averages for repeatability, intermediate precision, and the ratio between the two precision levels are mainly dependent on the type of drug product. Repeatabilities were found up to 0.8% for solutions, 1.6% for drug substances, 1.9% for tablets, 2.3% for creams, and 3.4% for a bath. For intermediate precision, which includes additional variability factors due to the reference standard, operator, equipment, reagents, etc., a similar dependency was obtained with a slightly changed order: up to 1.1% for drug substances, 2.2% for solutions, 2.3% for tablets, 3.1% for creams, and 3.2% for a bath. The ratio between the precision levels is up to 2.5 and similar for all investigated drug product types, apart from solutions with up to 5.3. These differences for the types of drug product may be explained by the influence of the sample and/or the sample preparation: the more complex, the higher the variability contribution. For the investigated examples, the impact of the analyte and of the concentration (dosage) seems to be of less importance. Therefore, a classification of drug product types for orientation on acceptable precision (ranges) for LC assay seems to be possible.

Keywords: Drug product; Drug substance; Analytical variability; Reproducibility; Analysis of variances; Regression

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

0731-7085/$ – see front matter © 2005 Elsevier B.V. All rights reserved.
doi:10.1016/j.jpba.2005.02.009

1. Introduction

Precision, i.e. "the closeness of agreement (degree of scatter) between a series of measurements" [1a], is of utmost importance for any analysis. In case of assay procedures, it determines directly their suitability because the analytical variability needs to be compatible to the acceptance limits of the specification and causes typically a large, if not the dominating, part of the specification range [2]. But also if these minimum requirements are fulfilled, in each validation the question of acceptance criteria is raised. Besides for the precision level itself, establishing an acceptable variability is crucial because many other performance parameters, e.g. in linearity and recovery, are linked with it [3].

For a proper interpretation and reasonable conclusions, it is important to address the precision level [1b] correctly. With respect to repeatability, it is essential to apply the whole analytical procedure (as described in the control test), not just to inject the same sample solution six times. This is also the reason to use authentic samples because only then the analytical procedure can be performed exactly as in the routine application. Intermediate precision includes the influence of additional random effects according to the intended use of the procedure in the same laboratory and can be regarded as an (initial) estimate for the long-term variability. Relevant factors, such as operator, instrument, and days, should be varied. Intermediate precision is obtained from several independent series of the analytical procedure (at least two) applied to preferably authentic and identical samples. In case of relative techniques, the preparation and analysis of the reference standard contribute significantly to the variability. Reproducibility is obtained from variations between laboratories.

In literature, usually individual repeatabilities and intermediate precision from only few series are reported. During stability studies, the same analytical procedure is applied over a long time. Therefore, these data are an excellent source to provide very reliable analytical variability [3]. If repeated determinations are performed for each storage interval, both overall repeatability and intermediate precision can be calculated, in case of sufficient replicates also individual repeatabilities. These precisions can be used in a specific as well as a general manner. The former can be regarded as part of the life-cycle concept of validation [4] for a particular analytical procedure, i.e. to accumulate information to obtain increasingly reliable estimates of the analytical variability. Collecting the obtained results provides the opportunity to define ranges to be expected for the precision levels and whether or not classifications are possible. Such classifications and expectations may serve as orientation for method development, establishing general validation acceptance criteria, or verifying the applicability of traditional specification limits [2,5]. For this purpose, the present investigation was started by the Working Group Drug Quality Control/Pharmaceutical Analytics of the German Pharmaceutical Society (DPhG) as a conclusion of the discussion on the consensus paper "Establishing Specification Acceptance Limits" [2,5].

2. Experimental

2915 assay values from 156 stability studies of 44 drug substances (DS) and drug products (DP) were compiled. The analytical procedures applied were typical reversed-phase LC assays with UV detection; the amount of analyte injected ranged from 0.125 to 40 μg, the concentration fraction of the analyte in the (original) sample (g/g) from 0.00042 to 100.0%. The analytes are mainly low-molecular weight synthetic drugs. As the objective of the investigation was to obtain precisions and their distribution for typical pharmaceutical applications, the analytes were not disclosed. The drug products were grouped into the following major types: aerosol (2), bath (3), cream (11), gel (4), lyophilisate (8), ointment (3), solution (24), suspension (2), and tablet (90). Drug substance (9) is regarded here as a drug product sample type too (the number in parentheses indicate the different stability studies). A prerequisite to calculate precision are non-rounded, individual results. In order to increase the number of replicates, several presentations, i.e. packaging variants, of the same bulk batch or several storage temperatures can be combined, provided they have the same stability and are analysed in the same series.

3. Calculations

All precisions are calculated and reported as relative standard deviations.

3.1. Individual repeatabilities

Individual repeatabilities were calculated from the assay values of independent sample preparations for each storage interval according to Eq. (1), if the number of values (n) was at least four. If different storage temperatures and packaging variants had no influence on the stability, the data were pooled. In case of two and three determinations per storage interval, the overall repeatabilities (Eqs. (2) and (3)) were included, as far as the overall degrees of freedom (d.f.) were less than 10. This restriction is supposed to avoid "too reliable" repeatabilities, because the aim was to investigate the distribution of individual precision results.

$$s_{i\%} = \frac{\sqrt{\sum (y_i - \bar{y})^2/(n-1)}}{\bar{y}} \times 100\% \quad (1)$$

3.2. Overall repeatabilities

Pooled repeatabilities over all storage intervals of a stability study were calculated according to Eqs. (2) and (3)

[6]. They were used to obtain the ratio between the precision levels in each study, additionally to represent (individual) repeatabilities if d.f. < 10 and to calculate average repeatabilities (of the respective subgroups), weighting the overall repeatabilities by their respective overall d.f. (Eq. (4)).

$$s_r^2 = \frac{\sum((n_j - 1)s_j^2)}{\sum n_j - k} \quad \text{or} \quad s_r^2 = \frac{\sum(s_j^2)}{k} \text{ (with equal } n) \tag{2}$$

$$s_{r\%} = \frac{\sqrt{s_r^2}}{\bar{\bar{y}}} \times 100\% \tag{3}$$

$$s_{av} = \sqrt{\frac{\sum(\text{d.f.}_{st} \times s_{r\%,R\%}^2)}{\sum \text{d.f.}_{st}}}, \quad \text{d.f.}_{st} = \sum n_j - k \tag{4}$$

n_j, s_j, $\bar{y}_j$ = Number of determinations, standard deviation, and mean for storage interval j; $\bar{\bar{y}}$ = overall mean of assay determinations for all storage intervals; k = number of storage intervals; d.f.$_{st}$ = overall degrees of freedom of a stability study.

3.3. Intermediate precision

In case of a statistically non-significant decrease in the content of the analyte, intermediate precisions are calculated by means of an analysis of variances (ANOVA, Eq. (6)) [6]. This precision level includes in addition to the variability within the storage intervals (Eq. (2)) the variance contribution between them, i.e. between the means of each storage interval (Eq. (5)). Homoscedasticity of the variances s_j^2 and a significant difference between the means $\bar{x}_j$, which, in a strict statistical sense, are prerequisites to proceed with the ANOVA calculations, were ignored in the present investigation because the objective was to compile experimental precision results as obtained in routine stability testing (see also [7]).

$$s_g^2 = \left(\frac{\sum(n_j \times \bar{y}_j^2) \times \sum n_j - \left(\sum(n_j \times \bar{y}_j)\right)^2}{(k-1) \times \sum n_j} - s_r^2 \right) \times \frac{(k-1) \times \sum n_j}{\left(\sum n_j\right)^2 - \sum(n_j^2)} \tag{5}$$

or

$$s_g^2 = \frac{\sum(\bar{y}_j - \bar{\bar{y}})^2}{k-1} - \frac{s_r^2}{n} \text{ (with equal } n) \tag{5a}$$

$$s_R^2 = s_r^2 + s_g^2 \quad \text{if } s_g^2 < 0: \quad s_R^2 = s_r^2, \quad s_{R\%} = \frac{\sqrt{s_R^2}}{\bar{\bar{y}}} \times 100\% \tag{6}$$

The equations used can be explained in the following way: The first term in Eq. (5a) describes the deviation of the individual means from the overall mean. Note that the variance within the series contributes to the spread of the individual means. Even if there were no real changes over time, i.e. no inter-series variance at all, the individual means would show spread. This variance contribution must be subtracted (second term in Eq. (5a)). Because the contribution is related to a mean, the variance of the individual determinations s_r^2 is reduced by a factor of $1/n$, i.e. it is scaled down to be comparable.

In addition to the aforementioned calculations, a simple overall relative standard deviation was calculated according to Eq. (1).

If degradation occurs, the inter-serial variance (Eq. (5)), obviously, is not anymore a measure of random variability, and consequently the ANOVA cannot be applied. In these cases, judged by the significance of the slope of the regression line obtained from all individual assay determinations (y-values) versus the storage time (x-values) (Eq. (9)), the residual standard deviation of the linear regression (Eq. (7)) was used. This parameter is a measure of the scatter of data around the regression line. For normalization, it is referred to the content mean (Eq. (8)), thus corresponding to the relative standard deviation of an intermediate precision. Within the shelf life of a pharmaceutical formulation, usually a linear degradation of the active ingredient is assumed [1c], justifying the application of a linear regression. Nevertheless, in this work several stability studies revealing a larger degradation (between 10 and 13%) were checked. The 95% confidence interval of the quadratic coefficient included zero in all these cases; thus, an alternative quadratic model does not provide a significant better fit of the data. Therefore, the assumption of a linear decrease is suitable for the use of these data to estimate precision. However, due to the weighing effect included in the normalisation, the decrease in the content should be limited to about 15%. Sometimes, significant upwards slopes were observed. As neither drug products nor packaging materials allowing for evaporation were included, the (statistically) significant increase of the content must be assumed to be of no practical relevance. Therefore, the calculation of the intermediate precision was here performed by an ANOVA.

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

$$s_{R\%} = \frac{s_y}{\bar{\bar{y}}} \times 100\% \tag{8}$$

$$\text{if } CI_b = t(95\%,\ n-2) \times \sqrt{\frac{{s_y}^2}{\sum(x_i - \bar{x})^2}} < b, \quad \text{the slope is significant} \tag{9}$$

y_i = individual assay of content at storage time x_i; a, b = intercept and slope of an unweighted linear regression line; $\bar{y}$, $\bar{x}$ = mean of all individual content determinations and

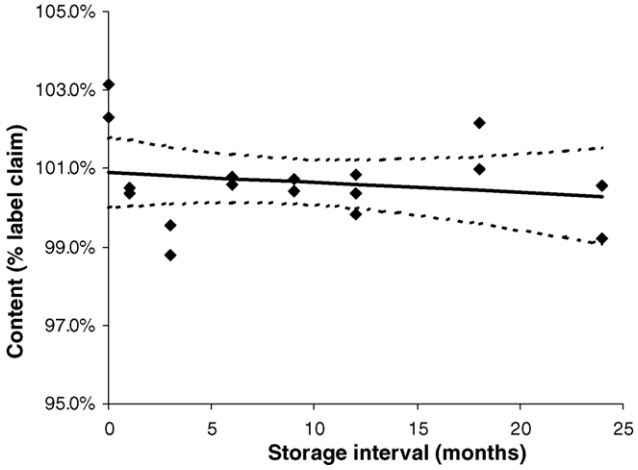

Fig. 1. Results of a stability study of a tablet stored at 25 °C/60% relative humidity. Unweighted linear regression results in an intercept and slope (with 95% confidence interval) of 100.89% and 0.025% (±0.074%), respectively. From the residual standard deviation, an intermediate precision of 1.10% is calculated. The regression line (solid) with its 95% confidence limits (dotted lines) is shown.

storage times; $t(95\%, n-2)$ = Student−t factor for 95% confidence level.

Average intermediate precisions of the respective subgroups weighted by the respective overall d.f. were calculated according to Eq. (4).

In Fig. 1, an example of a stability study of a tablet is shown. The confidence interval of the slope is larger than the slope itself, i.e. includes zero and consequently, the slope is not significant. Therefore, overall repeatability and intermediate precision can be calculated by addition of the variances after ANOVA (Eqs. (3) and (6)) resulting in 0.56 and 1.11%, respectively. Comparing the latter with the residual standard deviation of the regression of 1.10%, both calculation procedures result in identical intermediate precisions; even the overall relative standard deviation from all assay values, which is not the correct measure of the overall variability, is numerically very similar with 1.08%.

4. Results and discussion

4.1. General

The objective of this project was to extract a large amount of precision data from LC-assay stability studies in order to investigate their distribution ranges with sufficient reliability. Average parameters of specific analytical procedures can be regarded as estimates of the true values, for example, 0.63 and 0.89% for the repeatability and intermediate precision of a lyophilisate, respectively, (no. 10 and 11 in Figs. 2 and 3), or 1.03 and 1.75% for a tablet (no. 18 and 21). In contrast, averages for a combined group, such as LC assays of tablets, should be rather interpreted as orientation. However, these values serve their purpose as condensed information for the comparison of analytical methods and subgroups because they are less influenced by extreme results than the limits of the distribution ranges.

Both intermediate precision and reproducibility are measures of the long-term variability. Strictly, the latter is defined as between-laboratory precision and not essentially requested for submission of a new drug application [1a]. However, in the long-term perspective, it can be expected that the two sublevels approach each other, at least for applications within the same company. In the present investigation, the term intermediate precision was used, although it can be expected that during stability studies over up to 60 months (with an average of 28 months) in larger companies also different laboratories are included. In comparison to the intermediate precision ob-

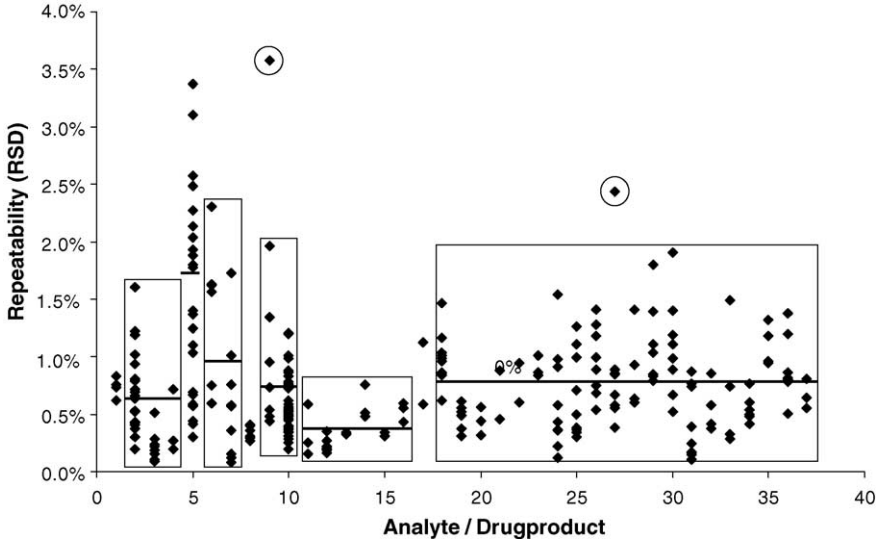

Fig. 2. Distribution of repeatabilities from stability studies. The number on the x-axis corresponds to the different analytes per drug product (for details, see Table 1). The distribution range and the average repeatability for the drug product types are indicated by rectangles and horizontal lines, respectively. Results excluded as probable outliers are shown in circles.

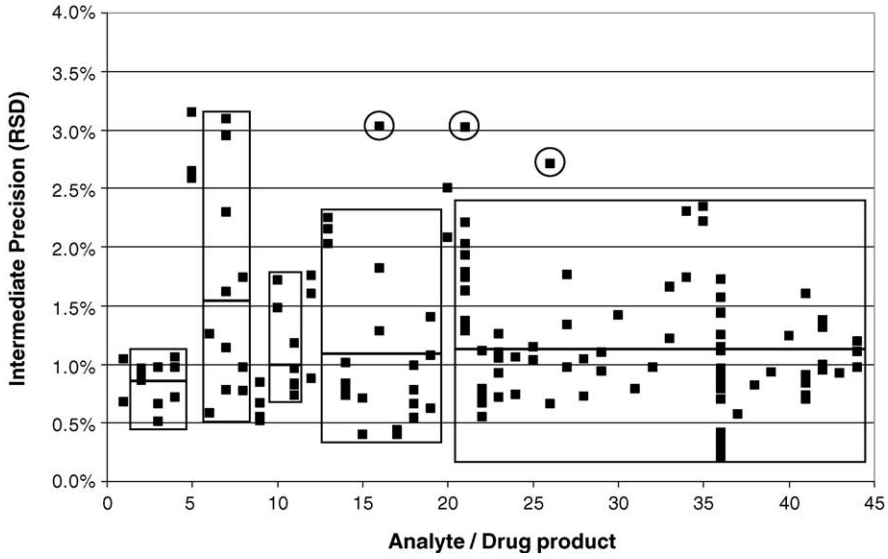

Fig. 3. Distribution of intermediate precisions from stability studies. The number on the *x*-axis corresponds to the different analytes per drug product (for details, see Table 2). The distribution range and the average repeatability for the drug product types are indicated by rectangles and horizontal lines, respectively. Results excluded as outliers are shown in circles.

tained for validation, where the number of series as well as the time frame is usually limited, the results from the presented investigation comprised between 4 and 11 storage intervals (i.e. independent series) with an average of seven. Therefore, the obtained intermediate precisions from stability can be expected to be more reliable, i.e. better estimates of the true variabilities. Reproducibility from collaborative trials can be expected to include additional contributions due to probably larger difference among the participating laboratories (such as equipment, experience with the product, "culture", etc.).

4.2. Repeatabilities

The reliability of standard deviations is strongly dependent on the number of values the calculation is based on. This can be illustrated by the upper limit of the 95% confidence interval [10]. For two, five, and nine d.f., the true standard deviation can be up to 4.4-, 2.1-, and 1.6-fold of the calculated result, respectively. It is obvious that a standard deviation calculated from three values only (unfortunately seen rather frequently in literature) does not provide meaningful information and should be avoided.

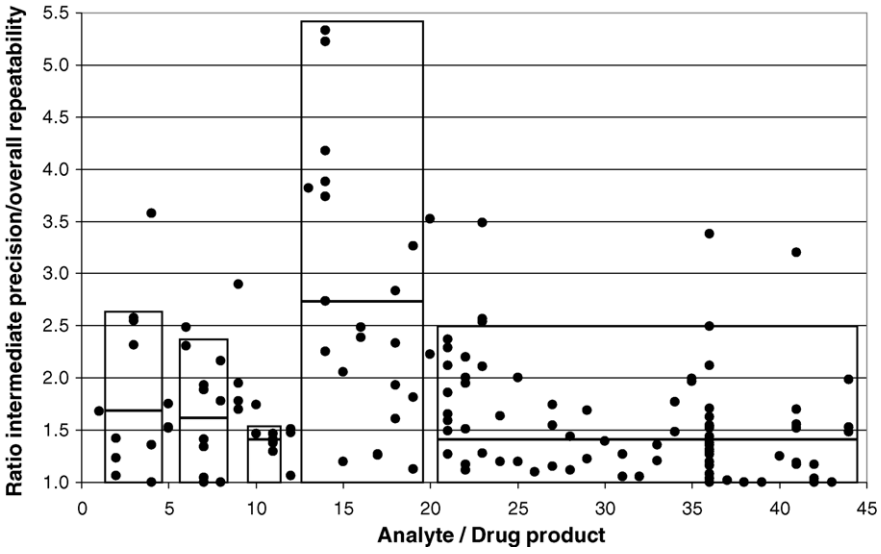

Fig. 4. Distribution of the ratios between intermediate precision and overall repeatability from stability studies. The number on the *x*-axis corresponds to the different analytes per drug product (for details, see Table 2). The 90% distribution limit and the average for the drug product types are indicated by rectangles and horizontal lines, respectively.

Table 1
Repeatabilities from stability studies

Drug product type (no. in Fig. 2)	No.[a]	Range (%)	Average[b] (%)	Ratio[c]
Aerosol (no. 1)	4	0.6–0.8	–	–
Drug substance (nos. 2–4)	32	0.1–1.6	0.64	2.5
Bath (no. 5)	24	0.3–3.4	1.73	2.0
Cream (nos. 6–7)	16	0.1–2.3	0.98	2.3
Gel (no. 8)	6	0.3–0.4	–	–
Lyophilisate (nos. 9–10)	44	0.2–2.0	0.72	2.8
Solution (nos. 11–16)	20	0.2–0.8	0.39	2.1
Suspension (no. 17)	2	0.6–1.1	–	–
Tablet (nos. 18–37)	120	0.1–1.9	0.80	2.4

[a] Number of repeatabilities.
[b] Calculated from overall repeatabilities according to Eq. (4).
[c] Between upper range limit and average.

Therefore, in order to investigate the distribution range of repeatabilities in the present study, four determinations were defined as a minimum number for individual repeatabilities (Eq. (1)), the maximum number available was eight. Overall repeatabilities (Eq. (3)) were included up to an overall d.f. of nine. The data are shown in Fig. 2 (individual results can be downloaded from the website of the DPhG Working Group, (http://www.pharmchem.tu-bs.de/waetzig-dphg_engl.html), grouped according to the type of sample (drug product).

The large distribution range of experimental variabilities obtained with the same analytical procedure for the same sample is obvious, especially when a larger number of results is available (Fig. 2, vertically arranged). For example, the repeatabilities for drug substance no. 2 scatter from 0.20 to 1.61%, (average 0.77%), for the lyophilisate no. 10 from 0.19 to 1.20% (average 0.63%), and for the tablet no. 25 from 0.30 to 1.26% (average 0.75%). Therefore, a limited amount of data, as usually obtained during validation, should be cautiously interpreted. It must also be taken into account that only the upper limit of the respective distribution is an appropriate acceptance criterion. As this information is not available for a particular method at the time of validation, generalisation and classification is required to provide orientation.

If the repeatability would be strictly analyte-specific, the same results should be obtained for all types of drug product with the same analyte. Nos. 1, 6 and 8, nos. 2 and 10, and nos. 3 and 29 in Fig. 2 represent the same analyte, but show clearly different repeatabilities. Instead, the variability seems to be rather dependent on the type of DP. The subgroups are indicated in Table 1 and illustrated in Fig. 2 by rectangles that also provide the limits of the distribution range. Two extreme values deviating from the other repeatabilities of the same DP type as well as more than three times from the corresponding average repeatability were excluded.

Average repeatabilities were calculated (Eq. (4)) from the overall repeatabilities to include all experimental results when enough data were available (Table 1). They can be regarded rather as "synthetic" summary parameters. Of course, the true repeatability for a specific analytical procedure can be expected to deviate from this average. However, this parameter can be used to provide a further estimate for the upper limit of the distribution, less influenced by exteme individual results. For most groups, the ratio between the upper limit and the average repeatability is rather similar between 2 and 2.5, which corresponds to the upper 95% confidence limit of a standard deviation calculated from five to six values.

Modelling of the variance distribution was also considered, if enough data were available, e.g. for tablets. Variances from a uniform random process are χ^2-distributed. Thus, we tried to fit different χ^2 functions, depending on the effective d.f. to our variances. On first glance, approximations with two or three d.f. gave a reasonable fit. However, comparing the confidence intervals obtained from the χ^2 functions to the empirical confidence intervals (data between the 5 and 95% quantile), the latter were certainly superior. As we know, the variances under investigation are not from a uniform random process. Apparently the χ^2 model functions are too sensitive to the violation of this assumption.

Taking both averages and upper range limits into account, solutions and the bath display the smallest and largest standard deviations. The other DP types are closer together in the sequence DS < lyophilisates ~ tablets < cream. Of course, the classification criteria may depend on both the specific procedures and analytes used in this investigation as well as on single results, especially the upper range limits. However, the same sequence results if the upper limit of 75% of all data is used (i.e. the "core-data" only). The only difference observed is in lyophilisates, which resemble more DS here. Therefore, it can be concluded that the distribution of the individual repeatabilities reflect the complexity of the sample and/or its preparation. For example, solutions usually require only dilution (if at all) with only minor variability contribution. In contrast, tablets are typically ground, weighed, dissolved or extracted, etc. Inhomogeneities of the sample or during sampling can also be expected to play a role. However, if of acceptable magnitude, this can be regarded as part of the routine analytical variability. Larger inhomogeneties caused by the manufacturing process are not regarded here because this is the objective of content uniformity testing. For batches used in stability studies, an appropriate content uniformity is guaranteed.

Repeatabilities obtained from literature are basically consistent with these results. The DS average corresponds very well with an overall repeatability of 0.6% from a collaborative trial of the European Pharmacopoia for the LC assay of cloxacillin [11,12]. For a combined group of investigated DS, lyophilisates, solutions and suspensions, an average and upper limit of 0.52 and 1.2%, respectively, was reported, for tablets 0.81 and 1.5% [3]. Other publications describe typical tablet repeatabilities of less than 2% (summarized in [8]). The large influence of the sample and/or preparation is confirmed by repeatabilities up to 5% for emulsions [13,14] and average and upper limit of 6.5 and 15%, respectively, for a chewing gum [15]. From the results of collaborative trials [16], average repeatabilities of 1.02, 0.85, and 1.42% for DS, solutions

and tablets can be calculated. Although the absolute values are larger than those obtained in this investigation, the order is the same. The improved results in this investigation may be attributed to the origin of the collaborative studies which was before 1985. In addition, larger variabilities must be expected in case of collaborative trials due to a lesser degree of familiarity with the analytical procedure compared to in-house methods applied in the present stability studies. This is also the reason for a thorough outlier testing in collaborative trials to avoid the inclusion of non-representative results [17]. Some discrepancies can be found with respect to solutions and DS, which are reported in literature up to 2.5 and 3%, respectively (summarized in [8]). This may be partly attributed to the fact that analytical procedures were sometimes optimised for the simultaneous determination of several analytes, but it can also be expected that some analytes/methods require larger variabilities (or optimization). Therefore, the average values and distribution ranges discussed should be regarded as orientation for typical applications.

4.3. Intermediate precision

The results for the intermediate precisions were calculated using a hierarchical design for the ANOVA analysis (3.3., Eqs. (4)–(6)) [6]. They are shown in Fig. 3 and Table 2. Note that three extreme values deviating from the other results of the same DP were excluded.

Compared to repeatability, a different order is obvious. For DS, the smallest intermediate precisions are observed, for creams and the bath the largest. Solutions, tablets, and lyophilisates display similar results, although for the latter a tighter distribution range seems to occur. However, this should be interpreted with caution because only eight results from two analytes are available. The sequence of the subgroups observed by investigation of averages and upper range limits is also confirmed by the upper 75% range of the distribution. Again, empirical intervals were used due to their advantages compared to confidence intervals obtained from χ^2 model functions (see Section 4.2). The change in the sequence can be explained by the different weight of the contribution of the variation factors in comparison to the repeatability. For solutions, the variability between the series (caused by reference standard preparation and analysis, equipment, operator, time, reagent effects, etc.) is much larger than the repeatability, resulting in (an average) variance contribution of only 13% for the repeatability. For DS and tablets, the corresponding contributions are 55 and 49%, respectively. Therefore, in case of less complex sample and/or sample preparation, the intermediate variation factors become more important for the overall variability. This is a further reason to abstain from applying statistical significance tests for homogeneity of variances and differences between the means in the ANOVA calculation of precisions. In case of validation, the preferred option would be to establish absolute upper acceptance limits for the various precision levels [9]. For this purpose, the present investigation can provide orientation for acceptable ranges of LC-assay precisions.

It is rather difficult to obtain and/or interpret intermediate precisions from literature. Usually, they originate from a smaller number of series (such as two). Therefore, these results are less reliable compared to the actual investigation where the number of storage intervals ranged from 4 to 11, with an average of seven. Intermediate precisions/reproducibilities from validation, transfer and some stability studies [3] were reported to have averages and upper limits, respectively, of 1.1 and 1.7% for DS, ∼0.7 and 1.3% for lyophilisates and solutions, and 1.4 and 2.3% for tablets. The results for the latter are similar, also in other investigations (summarized in [8]), whereas for DS lower and for solutions and lyophilisates higher variabilities are observed in the actual investigation. In collaborative trials, larger values were reported for reproducibility [16], with 1.5, 2.0, and 3.0% for DS, solutions, and tablets, respectively, but with the same ranking. This can be explained by the inclusion of additional variation factors between laboratories of different companies, compared to long-term applications in the same laboratory (company), as it was the case in the actual investigation.

Table 2
Intermediate precisions and ratios between the precision levels

Drug product type	No.[a]	Intermediate precision (%)		Ratio reproducibility/overall repeatability	
(no. in Fig. 3)		Range	Average	90% limit	Average
Aerosol (no. 1)	2/24	0.7–1.0	–	1.7[b]	–
Drug substance (nos. 2–4)	9/217	0.5–1.1	0.86	2.6	1.7
Bath (no. 5)	3/121	2.6–3.2	2.75	1.8[b]	1.6
Cream (nos. 6–8)	11/177	0.6–3.1	1.53	2.3	1.6
Gel (no. 9)	4/44	0.5–0.9	–	2.9[b]	–
Lyophilisate (nos. 10–11)	8/292	0.7–1.7	1.05	1.5	1.4
Ointment (no. 12)	3/48	0.9–1.8	–	1.5[b]	–
Solution (nos. 13–19)	20/390	0.4–2.2	1.10	5.3	2.7
Suspension (no. 20)	2/28	2.1–2.5	–	2.5[b]	–
Tablet (nos. 21–44)	90/1574	0.2–2.3	1.14	2.5[c]	1.4

[a] Number of intermediate precisions (corresponds to number of stability studies)/overall number of assay values.
[b] Largest ratio obtained.
[c] 95% limit.

The data from the 94 stability studies without a statistically significant decrease in the content were further investigated with respect to the suitability of a simple overall standard deviation and the residual standard deviation of the regression as measures of intermediate precision. The average ratio between the overall standard deviation and the ANOVA intermediate precision is 0.96 (±0.04). The average ratio with respect to the residual standard deviation is calculated to 0.95 (±0.06). Both ratios are close to unity. Therefore, it can be concluded that the residual standard deviation of the regression is a suitable measure for intermediate precision and its use in case of a significant (but limited) degradation is justified (see Section 3.3). The simple overall standard deviation can also be used as an estimate of intermediate precision if there is no change in the content during the investigation. However, as much more information, i.e. the various precision levels, can be obtained by application of an ANOVA at the same cost, this approach is preferable [7].

4.4. Ratio between intermediate precision and overall repeatability

This ratio corresponds to the difference between the two precision levels, i.e. the impact of the factors varied in long-term application. A classification of these factors would allow a prediction of the long-term variability from repeatability determinations, which are more readily and easily available. 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. This, of course, does not reflect the real situation because in case of relative methods, at least the contribution from the reference standard (i.e. $\sqrt{2}$), and most likely additional effects can be expected. However, this can occur experimentally when one or several experimental repeatabilities are obtained in the upper range of the distribution, thus covering the differences between the series. It must be taken into consideration that the uncertainty of these ratios is larger because it includes the uncertainty of both precision levels. Therefore, in order to minimize the influence of extreme results, the upper 10% were excluded for estimating the limit of the distribution and for calculating the average of the ratios. Due to the larger number of ratios available for tablets, the 95% distribution limit was taken. An upper limit of 2.5 and an average of 1.5 were observed for all types of DP, apart from solutions (see Fig. 4). There, markedly larger ratios up to 5.3 with an average of 2.7 were found. The larger the repeatability for a given group of samples, the smaller is the weight of the additional variability contributions for intermediate precision, such as reference standard preparation and analysis, operator, time, etc. Consequently, the ratio is also smaller, and vice versa. From the ratio, the error contribution of repeatability to the overall variability can be directly calculated as the squared reciprocal (variance of repeatability/variance of intermediate precision). For example, in case of solutions, the larger ratio may be explained by the simple sample preparation, resulting in a repeatability contribution of only 14% (using the average ratio of 2.7). As a consequence, the influence of the reference standard and the other variations to the overall variability is increased, affecting directly the intermediate precision. In contrast, for more complex samples as the bath, the repeatability is dominating, resulting in small ratios. In accordance, for emulsions [14] and chewing gum [15] ratios of 1.1 to 1.2, respectively, are reported. The distribution limit of the ratios for lyophilisates seems to be smaller but this may be caused by the low number of data available. Therefore, it is not possible to evalutate their representativity.

These findings are in agreement with the more general estimation of factors between the precision levels of about 1.5 per level [18], i.e. a ratio of 2.2 for repeatability and long-term precision.

4.5. Concentration dependency

Processing a large number of data from collaborative trials with various analytes, matrices, and analytical techniques over large concentration ranges, Horwitz et al. found a striking simple exponential relationship between the relative standard deviation among laboratories, i.e. reproducibility, and the mass fraction of the analyte (Eq. (10)). The standard deviation decreases less rapidly than the concentration, resulting in an increase of the relative standard deviation for lower concentrations. The Horwitz relationship is used as a benchmark for the performance of laboratories in collaborative studies, with acceptable reproducibility "within one-half to twice the value predicted by the equation from the concentration" and repeatability between "one-half to two-thirds the among-laboratory variability" [19].

$$\mathrm{RSD}_{R,\mathrm{predicted}} = 2 \times C^{-0.1505} \tag{10}$$

In an investigation of HPLC collaborative studies, a linear relationship was found for the plot of the reproducibility versus the logarithm of the concentration. The precision increased about 0.4% for each 10-fold decrease in concentration, from 2% for a 100% concentration to 3.6% at a concentration of 0.01% [16].

The present intermediate precision and repeatability data were investigated, both with respect to the concentration fraction, i.e. the dosage, and the amount of analyte injected. For intermediate precision, a significant linear correlation was found with p-values of 7×10^{-5} and 2×10^{-5} for concentration fraction (Fig. 5) and amount injected (Fig. 6), respectively. However, the limits of the distribution range do not show a comparable concentration dependency. They display a plateau at about 3% intermediate precision below 1% concentration fraction or below 0.6 μg analyte injected. The value of the intermediate precision increases less rapidly than predicted for the reproducibility by the Horwitz equation (Fig. 5), even if the factor (originally 2, corresponding to 2% reproducibility at 100% concentration) (see Eq. (10)) was corrected for the 100% result calculated from the linear regres-

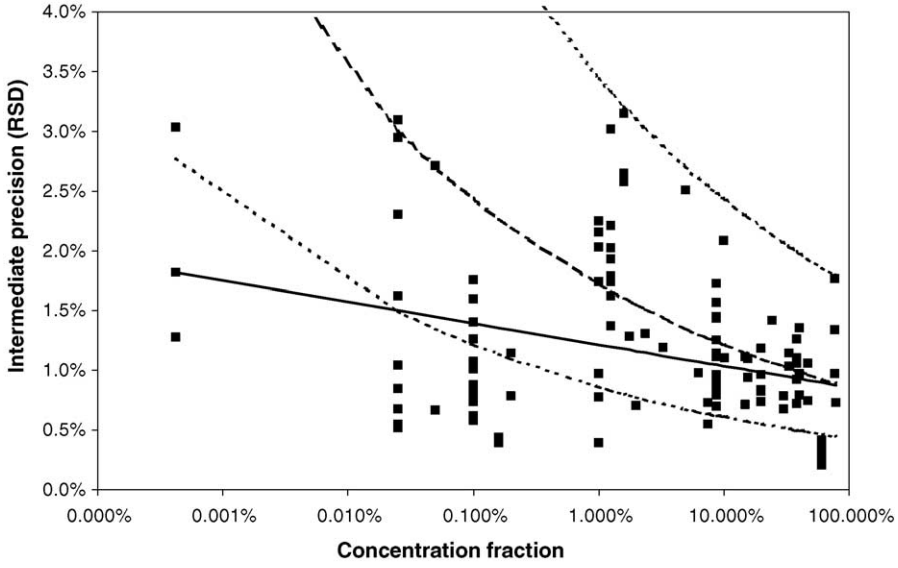

Fig. 5. Linear regression of intermediate precision versus the logarithm of the concentration fraction of the analyte in the dosage form. Besides the regression line (solid), the predicted Horwitz reproducibility (broken line) with the upper and lower expectation limit (dotted lines) are displayed.

sion. For repeatability, no significant linear correlation was observed (*p*-values of 0.73 and 0.09 for amount injected and concentration fraction, respectively).

The only marginal concentration dependency in the present investigation supports the proposal that for LC assays, the drug product type, i.e. the sample and/or sample preparation, determines primarily the analytical variability. The observed significance for the concentration trend is not surprising because of the high number of data included in the regression. With 141 values, a coefficient of correlation of only 0.166 becomes already significant. The small trend observed can also be explained by the leverage effect of the DS samples. Their intermediate precisions are the least of all sample types, and their concentration fraction is close to 100%. As all other types of drug products display higher variabilities and lower concentration fractions, a trend is easily obtained.

Several explanations can be found for the observed deviation from the Horwitz relationship. First, it describes a general concentration dependency of precision within large concentration ranges for a multitude of analytical techniques. Second, it concerns reproducibilities obtained from collaborative studies, where, as already discussed, additional variability effects can be expected, which become probably larger for very small concentrations due to more complex sample preparation and matrix interferences. It could also be shown that the variability increases rapidly if the quantitation limit is approached because the integration error becomes the dom-

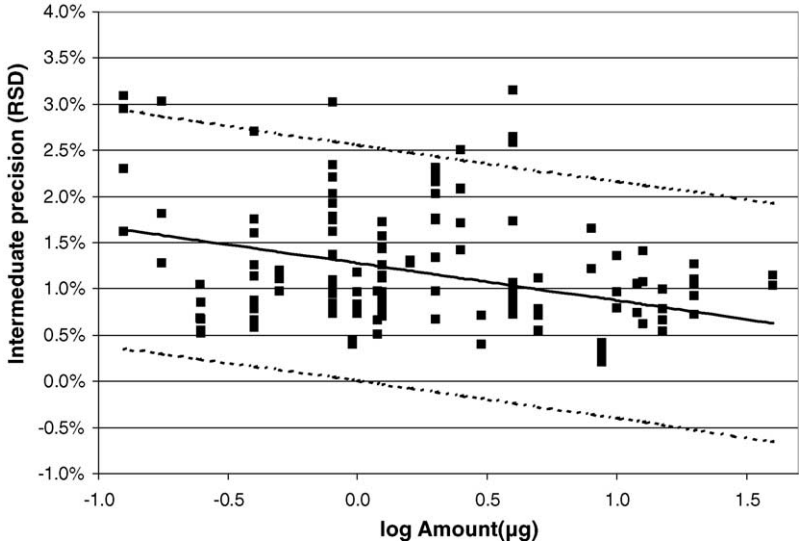

Fig. 6. Linear regression of intermediate precision versus the logarithm of the amount of analyte injected. The regression line (solid) and the limits of the 95% prediction interval (broken lines) are displayed.

inating variability contribution [20]. In the present investigation, the conditions are well controlled. Only assay procedures are included with concentrations well above the quantitation limit. Therefore, a closer look at the behaviour of the precisions became possible.

5. Conclusions and outlook

In this investigation, the precisions of LC assays are mainly influenced by the type of drug product, only to a minor extent by the concentration. The distribution of individual repeatabilities reflect the complexity of the sample and/or its preparation, i.e. up to 0.8% for solutions, 1.6% for drug substances, 1.9% for tablets, 2.3% for creams, and 3.4% for a bath were found. For intermediate precision, which includes additional variability factors due to the reference standard, operator, equipment, reagents, etc., a similar dependency was obtained with slightly changed order: up to 1.1% for drug substances, 2.2% for solutions, 2.3% for tablets, 3.1% for creams, and 3.2% for a bath. The change in the order can be explained by the larger importance of the intermediate variability contribution (inter-group variance) to the overall precision for less complex drug products with small repeatabilities. This is also reflected by a larger factor between the precision levels, which is up to 5.3 for solutions and up to 2.5 for the other types investigated.

The distribution intervals obtained in this investigation should be regarded as orientation. For most of the drug product types, apart from tablets, the number of results is (still) not large enough to allow a confident generalization. Therefore, we like to encourage all analysts to perform investigations as described for their products and analytes. An Excel file for calculation of precisions can be downloaded from the website of the DPhG Working Group (http://www.pharmchem.tu-bs.de/waetzig-dphg_engl.html). There is also a template for compiling precision data available, which can be sent to the chairman of the Working Group, Prof. Hermann Wätzig (h.waetzig@tu-bs.de). These results will be appropriately published in order to increase the reliability of the conclusions and to extend, if possible, to other dosage forms and techniques.

Acknowledgements

Stability studies that are the basis of the presented investigation require a large effort with respect to logistics, analytical activities, as well as evaluation. Therefore, we like to acknowledge the skilled work and contribution of the many of our colleagues involved in these studies.

References

[1] (a) ICH: Q2A, Text on Validation of Analytical Procedures, 1994;
(b) ICH: Q2B, Validation of Analytical Procedures: Methodology, 1996;
(c) ICH: Q1E, Evaluation of Stability Data, 2002.
[2] J. Ermer, H. Wätzig, Regul. Aff. J. (2003) 1–3.
[3] J.Ermer, H.-J. Ploss, J. Pharm. Biomed. Anal., 2004, in press.
[4] G.C. Hokanson:, Pharm. Technol. 18 (1994) 92–100.
[5] H.Wätzig and J.Ermer. Establishing Specification Acceptance Limits. Consensus Paper of the Working Group Drug Quality Control/Pharmaceutical Analytics of the German Pharmaceutical Society (DPhG). 2003. http://www.pharmchem.tu-bs.de/dphg_pospapier_eng03.pdf.
[6] 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.
[7] USP 28-NFZ3, General Chapter (1010), 2005.
[8] J. Ermer, in: J. Ermer, J.H. McB.Miller (Eds.), Method Validation in Pharmaceutical Analysis, Chapter 2, Wiley VCH, Weinheim, 2005.
[9] MVA – Method Validation in Analytics. 2001. NOVIA GmbH, Frankfurt, Germany. http://www.novia.de.
[10] R.M. Bethea, B.S. Duran, T.L. Boullion, Statistical Methods For Engineers and Scientists, Marcel Dekker, New York, 1995, 137.
[11] A.G.J. Daas, J.H.M. Miller, Pharmeuropa 9 (1997) 148–156.
[12] A.G.J. Daas, J.H.M. Miller, Pharmeuropa 10 (1998) 137–146.
[13] I.P. Nnane, L.A. Damani, A.J. Hutt, Chromatographia 48 (1998) 797–802.
[14] D. Castro, M.A. Moreno, J.L. Lastres, J. Pharm. Biomed. Anal. 26 (2001) 563–572.
[15] A.I. Gasco-Lopez, R. Izquirdo-Hornillos, A. Jiminez, J. Chromatogr. A 775 (1997) 179–185.
[16] W. Horwitz, R. Albert, J. Assoc. Off. Anal. Chem. 68 (1985) 191–198.
[17] W. Horwitz, J. Assoc. Off. Anal. Chem. 60 (1977) 1355–1363.
[18] B. Renger, J. Chromatogr. B 745 (2000) 167–176.
[19] R. Albert, W. Horwitz, Anal. Chem. 69 (1997) 789–790.
[20] U. Schepers, J. Ermer, L. Preu, H. Wätzig, J. Chromatogr. B 810 (2004) 111–118.

Further reading

[21] http://www.fda.gov/cder/ode4/preind/Glossary.htm (05-27-1999).
[22] http://encyclopedia.thefreedictionary.com.
[23] ICH Q6A, Specifications: Test Procedures and Acceptance Criteria for New Drug Substances and New Drug Products, Chemical Substances, October 1999.

Glossary

Analytical variability: variability induced by unmeant (random) changes in the analytical procedure

Assay: a quantitative measurement of the analyte in a given sample, i.e. the major component(s) in a drug substance or these component(s) (active ingredients) in a drug product [1a]

Bulk batch: batch of the respective pharmaceutical dosage form (e.g. tablets, capsules) before the packaging step

Confidence interval: interval around a calculated statistical parameter in which the true value is located with a determined statistical probability

Drug Product (DP): a finished dosage form, for example, tablet, capsule, or solution, that contains a drug substance, generally, but not necessarily, in association with one or more other ingredients [21]

DP: see drug product

Drug substance (DS): an active ingredient that is intended to furnish pharmacological activity or other direct effect in the diagnosis, cure,

mitigation, treatment, or prevention of disease or to affect the structure or any function of the human body [21]

DS: see drug substance

Homoscedasticity: a sequence of random variables is homoscedastic if all random variables in this sequence have the same finite variance [22]

ICH: International Conference on Harmonisation of technical requirements for registration of pharmaceuticals for human use

Intermediate precision: day-to-day or between-run precision, expresses within laboratories variations, e.g. different days, different analysts, different equipment [1a]. This precision level includes, additionally to the random variability of the measurement, the influence of the reference standard and of external factors (e.g. temperature, humidity, quality of reagents, operators' qualification etc.)

Inter-serial variance: variance between different series

Precision: the closeness of agreement (degree of scatter) between a series of measurements obtained from multiple sampling of the same homogenous sample under the prescribed conditions [1a]

Range: interval (between upper and lower limits) where parameters are found, can be expected, or allowed to occur

Repeatability: the precision under the same operating conditions over a short interval of time, intra-assay precision [1a]

Reproducibility: expresses the precision between laboratories, e.g. collaborative studies [1a]

Specification: list of tests, references to analytical procedures, and appropriate acceptance criteria, which are numerical limits, ranges, or other criteria for the tests described. [23]

Validation: process of proving that a method is suitable for its intended purpose [1a]

Variance: the square of standard deviation as a degree of precision