Laboratory Management   **Edited by: E. Wieland**

Christian Beier*

Recommended changes of the current version of the German Rili-BAEK

Abstract: A number of improvement proposals and corrections of the German Rili-BAEK (Guideline of the German Medical Association on Quality Assurance in Medical Laboratory Examinations) are discussed with special focus on the internal and external quality assurance (IQA/EQA) as well as reference intervals for quantitative results. Particular attention is paid to reconsider the retrospective analysis of control measurements. Such an analysis can be very useful to monitor establishing errors of measurement even before they become critical. The present method "Quadratischer Mittelwert der Messabweichung (QMMA)" has proved to be ineffective. Furthermore, the current idea of a common limit for single control measures and the retrospective statistics must be revised. As a more sophisticated concept, the novel Adaptive Retrospective Monitoring (ARM) has been developed. ARM is recommended as the new minimum requirement for the entire internal quality assurance. Further proposals to enhance clarity are given concerning the release decisions of medical devices and the EQA. Individualized medicine begins with a patient-specific interpretation of analytic results. This requires standardized subgroup-specific reference intervals with smooth age-related adaptations. Only large laboratories are able to ensure the desired specificity and a sufficient statistical significance of self-developed in-laboratory reference intervals. Hence, the need of a central database for harmonized reference intervals is discussed and recommended. Suitable and consistent reference intervals are also an essential prerequisite for unitless laboratory values like the zlog value.

*Correspondence: Dr. rer nat Christian Beier, ADMEDIA MVZ Waldenburg, Weg des Friedens 4, 08396 Waldenburg, Germany, E-Mail: christian.beier@admedia-diagnostik.de

Brief summary: A number of improvement proposals and corrections of the German Rili-BAEK are discussed with special focus on the internal and external quality assurance as well as reference intervals. The Adaptive Retrospective Monitoring has been newly developed to solve problems of the current mandatory strategy for internal quality assurance. A central database of harmonized, reliable and subgroup-specific reference intervals is strongly recommended.

Introduction

Since the early 1970s, the Guideline of the German Medical Association on Quality Assurance in Medical Laboratory Examinations – Rili-BAEK [1, 2] contains all legally binding regulations to run a medical laboratory in Germany. Besides management policies, it particularly includes detailed regulations concerning the quality assurance of measuring processes. As one of the most extensive and highly developed guideline, the Rili-BAEK serves as a reference for other national directives for clinical chemistry and laboratory management. The increasing complexity of the guideline as well as the huge technological progress in clinical chemistry require regular revisions of the Rili-BAEK. Last major releases were published in 2001 [3], 2007/2008 [4] and 2014 [1]. A new revision is already planned.

The present regulations of the internal quality assurance prescribe single measurements of a control sample (SMC) and a statistical retrospective analysis (RA) of quantitative SMC data at the end of an evaluation period. The German regulations solely accept control samples with a known target result determined by a reference institution (reference method value) or the manufacturer (nominal value). Thus, using such control samples, the internal quality assurance is able to evaluate both the imprecision and bias (i.e. inaccuracy) with regard to a predefined target value of each analyte on each platform in use. The Rili-BAEK exclusively defines

a maximum permissible limit for the relative total error Δ^{rel} (in percent, table B1a–c, column 3) [1], which is an Euclidean combination of imprecision and bias. The absolute value of the maximum permissible total error is given by

$$\Delta_{max} = \frac{\Delta^{rel} \cdot y_0}{100\%} \text{ and } \Delta_{max} = \sqrt{\frac{n-1}{n}s_{max}^2 + \delta_{max}^2} \qquad (1)$$

(y_0: target value of the control sample, s_{max}: maximum permissible imprecision, δ_{max}: maximum permissible bias). The factor $(n-1)/n$ can be neglected for sufficiently large n. For all analytes not listed in table B1a–c, the evaluation of internal laboratory deviation limits is required. These limits are revealed by an Euclidean sum of the bias and the k-fold imprecision obtained during an evaluation period (k is usually set to 3). Instead, the deviation limits provided by the manufacturer are valid, if they are more restrictive or the evaluation/update of own limits is still in progress.

The defined minimum requirements for SMC imply the measure of a single control sample per day of use. In cases of continuous operation exceeding 16 h, restarting, recalibration or every other intervention into the measuring system, a further measure of a control sample has to be performed. If at least two admitted control samples with different known analyte concentrations are available, two control samples have to be used in an alternating series. Nevertheless, manufacturer's instructions usually prescribe to measure both control samples in parallel. Maximum permissible limits for the total deviation of measurement have to be defined for all locally examined analytes. These limits must at least be consistent with the Rili-BAEK, table B1a–c, column 3 and the present manufacturer's specifications.

During consecutive evaluation periods with intervals of 1–3 months, SMC data have to be collected for any control sample in use. Each evaluation period is finalized by a RA of the recorded SMC data. If less than 15 values have been recorded within the first month, the evaluation period is prolonged by 1 month. Amounts of SMC values per evaluation period usually lies between 15 and 31.

The presented publication gives several recommendations for revisions of the internal and external quality assurance (IQA/EQA). Special emphasis is given to provide a more efficient statistical analysis of recent measurements of control samples to early identify a deterioration of the measuring performance. Furthermore, the constitution of a central database for reliable and more individualized reference intervals is recommended, to prepare for future achievements in clinical chemistry.

Distinct rules for release decisions of analytic devices

The present version of the Rili-BAEK should provide more objective and clearer conditions necessary to release the measuring system after yielding an off-limit SMC. A second measure of an ideally fresh control sample is not explicitly mentioned, although it is a useful and popular action, due to the fact that a contaminated or aged control sample is a common reason for off-limit values.

Actually, two alternate procedures exist, if an obtained SMC value exceeds a given limit. Both are not in contradiction to the present Rili-BAEK. First, a release is finally left to the decision of the responsible person without the explicit need to repeat the SMC. This could be justified by medical urgency, although a regular practice is questionable. The "positive" point of such a subjective decision is that the finally accepted SMC value is off-limit, which is a prerequisite for the current Quadratischer Mittelwert der Messabweichung (QMMA)-based RA to indicate out-of-control situations (discussed later). In this case, the policies of the Rili-BAEK should at least be extended to call for logged visual inspections. However, a visual inspection can never fully justify that the measured value is treated as a single outlier. This procedure should be clearly limited to sporadic and non-consecutive events in SMC determination.

The second approach requires a rough initial check of the measuring system in response to an outlier. Before turning toward a prolonged inspection in detail, it permits one repeated measure of the control sample (taking a fresh aliquot if possible). The number of permitted repeats has to be strictly limited to one. If the SMC fails again, detailed checks are unavoidable. Using the current QMMA-based RA in combination with this second variant, all measured SMC values (particularly including outliers) need to be accounted. Only if an outlier is unequivocally dedicated to an obvious momentary mistake or a compromised control sample, this particular outlier should be rejected.

It appears necessary to provide more detailed objective conditions for the release of measuring processes. The secondly described variant is favored for non-urgent situations. It may also be conceivable to use two-stage deviation limits. Exceeding the lower limit requires a logged inspection of the procedure and device(s) but without an obligation to repeat the SMC. If no problem could be identified, the analytical device is once released for regular use. Exceeding the lower limit would also be a clear indication to calibrate and/or exchange the control material or reagents before the next regular SMC. Exceeding the higher limit (e.g. 1.5-times higher or even more) requires full inspection,

cleaning, possible exchange of reagents/control material and finally a second SMC. The second of two consecutive measures, both located between lower and higher limits, has to be treated as being above the higher limit.

The misleading term "QMMA"

Currently, the RA facilitates a statistical method, which is denoted as "*Quadratischer Mittelwert der Messabweichung (QMMA)*" [1] (direct translation: squared mean of measuring dispersion). This phrase is problematic due to inconsistencies compared to the corresponding mathematical equation:

$$\text{QMMA:} \quad \Delta = \sqrt{\frac{1}{n} \cdot \sum_{i=1}^{n}(y_i - y_0)^2} \tag{2}$$

(n: trajectory length, y_i: single trajectory values, y_0: target value).

The equation is obviously a square root. Thus, Δ is a deviation rather than a variance as implied by the text phrase. Furthermore, the correct sequence of the applied mathematical operations would be given by the term "mean of squared differences" in contrast to a "squared mean of differences" that is suggested by the present phrase "*quadratischer Mittelwert*".

The discrepancies between formula and text phrase are significant; hence, a correction is recommended to meet the high quality of the guideline. Unfortunately, the term "QMMA" is widely used since many years. A pragmatic solution would be to use an anglicism in future. The term "QMMA" could, however, still be tolerated for historic reasons. The English phrase of the used eq. (2) is perfectly short and plain. The "root mean square deviation" (RMSD) also indicates the correct operational sequence. However, the term RMSD is already associated with the empirical standard deviation. To indicate the contrary use of a target value instead of the mean, the phrase "root mean square total (or target) deviation" (RMSTD) is finally recommended. Anyway, a new method will be presented below, which is intended to replace the existing RMSTD-based RA, due to clear issues of the present approach.

Remarks about the retrospective analysis

The RA of single measures of the control sample (SMC data) at the end of an evaluation period is regulated in

Rili-BAEK, part B1, chapter 2.1.3, paragraph 1 as follows: "*Aus den Ergebnissen aller Kontrollprobeneinzelmessungen, die zur Freigabe des Messverfahrens oder der Patientenergebnisse geführt haben, ist nach Beendigung der Kontrollperiode unverzüglich der relative quadratische Mittelwert der Messabweichung zu errechnen…*" [1] (translation according to [2]: Based on the results from all single measurements of control samples that have led to the release of the measuring procedure or of the patient results, the relative root mean square of the deviation of measurement has to be calculated immediately after the completion of an evaluation period…).

The explicit limitation to SMC data that have (directly) led to the release of the measuring procedure is initially astonishing. Although not explicitly claimed in the current version of the Rili-BAEK, one may think that only SMC values within their related deviation limits can finally permit the release of an analytical device. Thus, after measuring an off-limit SMC value, it is common practice to analyze the control material a second time (e.g. with a fresh control sample) just after a rough inspection of the equipment. Frequently, the second value is again in control. The final release decision is then achieved due to this second SMC. In this case, the paragraph of the Rili-BAEK clearly intends to ignore the previous off-limit value while calculating the RA. However, it has to be mentioned that the same maximum deviation limits are applied to both the particular SMC as well as the result of the RA. In the case that only those SMC data are mentioned that lied within the deviation limits (no outliers), the final value of the RA is inevitably within these deviation limits as well. Thus, the present implementation of RA makes no sense, unless the measuring procedure was released based on an off-limit (but nevertheless accepted) SMC value. Moreover, the RA can only fail, if such off-limit SMC values occurred in a significant amount during the evaluation period. Such a rather arbitrary concept of release decisions, tolerating off-limit SMC values without clear directives, should stay in conflict with the general intention of the Rili-BAEK. However, subjective release decisions, even in situations without a clear medical necessity, do not contradict to any present rule of the guideline.

Occasional outliers are even intended in the reference publication of the present mandatory RA method [5]. Thus, it is at least necessary to consider all measured SMC data including each outlier. However, exceptional situations (like an obvious temporary mistake or a compromised control sample) might exist, which justify to reject a particular outlier. Unfortunately, many of the current maximum permissible deviation limits defined in the Rili-BAEK seem to be too high to generate sufficient amounts

of outliers, necessary for an efficient application of the present RA. Clarification is required regarding the correct use of SMC data for RA. Further, the use of common maximum permissible limits for RA and single SMC values will be revised in a separate chapter.

An evaluation period should ideally be finished within about 1 month. Time periods of 2 months or more are often too long for beneficial post-facto corrective actions. Wrong therapeutic decisions might have already led to irreversibly adverse health effects on patients. In common situations where just one SMC value per day and control sample is generated, the minimum amount of 15 values per sample cannot regularly be achieved within 1 month, due to potential downtimes. Nevertheless, at least 15 values are indeed essential for a useful statistical analysis. A better solution would be to prolong evaluation periods by 1- or 2-week increments until 15 or more values are collected. Evaluation periods may also overlap if monthly cycles are generally preferred.

Common vs. different limits for SMC and RA

Since the Rili-BAEK version of 2008, a sole maximum permissible limit (table B1a–c, column 3) is intended for both the statistical RA by RMSTD/QMMA and each single SMC result. This decision seems to originate on an explanatory statement by Mcdonald [5] in 2006. Although his publication provides a significant contribution to the German quality assurance, the reasoning that leads to a common limit for SMC and RMSTD/QMMA has to be rejected. Contrary to his derivation, eq. (9) in [5], given as

$$\left| y_i - y_0 \right| \leq 2 \cdot s_{max} + \delta_{max} \qquad (3)$$

(y_i: SMC value, y_0: target value, s_{max}: maximum permissible imprecision, δ_{max}: maximum permissible bias), cannot be implemented into eq. (5) in [5]. His eq. (5) is equivalent to eq. (2) here. At $n = 1$, the statistical value Δ given in eq. (2) cannot be set equal to any arbitrary single-measure deviation $|y_i - y_0|$. It still represents the theoretical mean deviation $\Delta_{n=1} = < | y_i - y_0 |>$ only. Therefore, the permissible limit for SMC deviations $| y_i - y_0 |$ [eq. (3)] is not similar to the average deviation $< | y_i - y_0 |>$ or to the permissible limit of Δ [as given in eq. (1)].

In addition, Figure 1 illustrates two simple examples of normal-distributed SMC trajectories with 30 data points each. Both have an RMSTD/QMMA value Δ, which meets the common limit Δ_{max} [see eqs. (1)+(2); hereafter
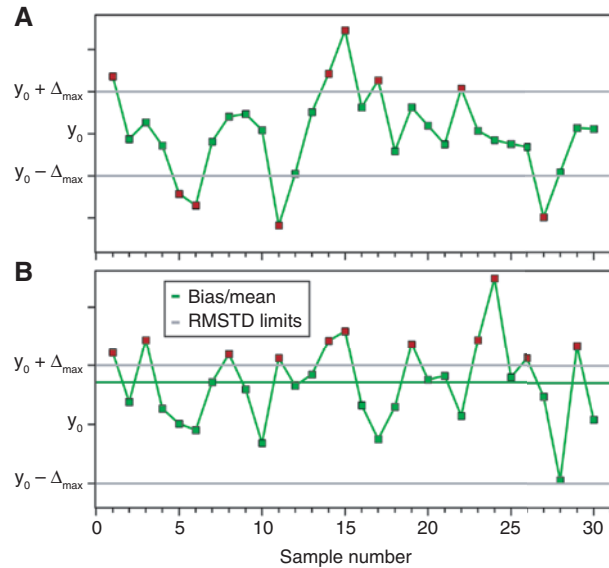

Figure 1: Exemplary trajectories of SMC data with normal-distributed deviations and a zero or constant bias with respect to the target value y_0.
(A) Trajectory with zero bias and maximum imprecision $s = \Delta_{max}$; (B) trajectory with equal bias and imprecision $\delta = s = 0.707 \cdot \Delta_{max}$ (bias level in dark green). Both RMSTD values are equal to the given maximum permissible limit Δ_{max}, thus the RA indicates a narrowly in-control situation. The limits $\pm \Delta_{max}$ are given as gray lines. Although statistically in control, both trajectories show several single off-limit values marked in red.

$\Delta_{max} = L_{RMSTD}$]. Thus, the RA of both trajectories indicates narrowly in-control situations. The trajectory in Figure 1A shows no bias and a dispersion equal to Δ_{max} and that in Figure 1B represents a 1:1 ratio of bias and imprecision ($s = \delta = 0.707 \cdot \Delta_{max}$). Although statistically in control, a significant number of SMC values indeed exceed the common limit. There are nine (expected 32%) in Figure 1A and 11 (expected 35%) in Figure 1B off-limit values.

Thus, the limit of an RMSTD/QMMA-based statistical RA (L_{RMSTD}) has essentially to be more restrictive than the limit for particular SMC deviations (L_{SMC}). If a common limit is used, one has either to deal with a number of outliers (limit too strict for SMC) or the RA is inefficient or even useless (limit too tolerant for RMSTD). One can define a relation factor λ between optimal choices of both limits

$$\lambda = \frac{L_{SMC}}{L_{RMSTD}} = \frac{\kappa \cdot s_{max} + \delta_{max}}{\sqrt{s_{max}^2 + \delta_{max}^2}}, \qquad (4)$$

where κ is an expansion factor with regard to the desired confidence level. Unfortunately, the factor λ depends on the ratio between δ_{max} and s_{max}. Figure 2C illustrates the mathematical relation between λ and δ_{max}/s_{max} for three common κ. To get a profound idea about the ratio
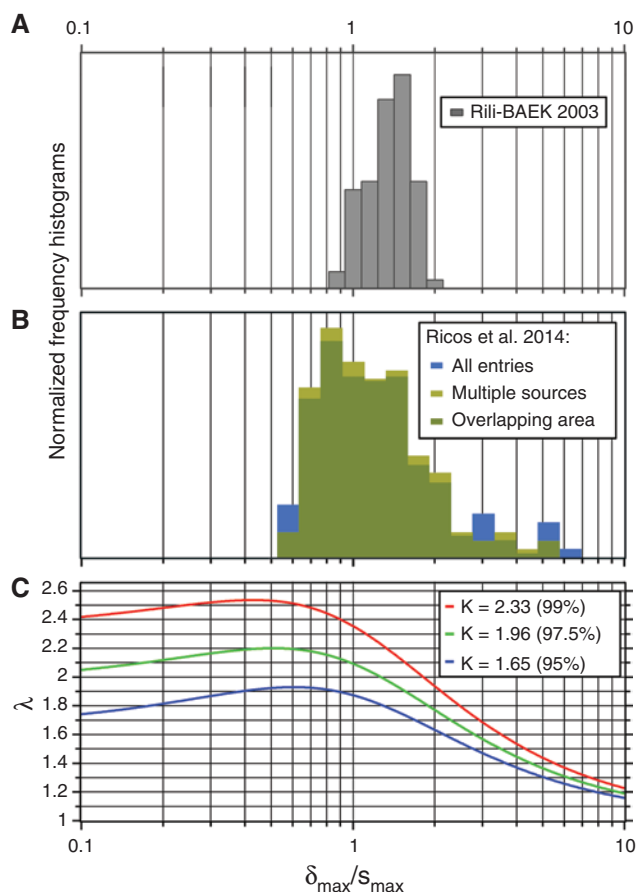
Figure 2: Relation between maximum permissible bias and imprecision of several analytes needed to approximate λ. (A) and (B) Frequency histograms of the ratios of maximum permissible bias vs. imprecision of medical analytes. The ratios are presented on a $\log_{10}$ scale. As referenced in the main text, the data sources are (A) Rili-BAEK 2003 and (B) Ricos et al. 2014. The histogram marked as "multiple sources" only considers data based on multiple reference sources according to column 3 [6]. Sample sizes and ratio averages are: Rili-BAEK: 90, 1.34; Ricos all: ~290, 1.21; Ricos multi: 150, 1.1. Average determination and histogram sampling were done in $\log_{10}$ space to provide almost equidistant interval lengths. (C) Functional relation between λ [see eq. (4)] and the maximum permissible bias/imprecision ratio for three different κ. Related one-sided confidence levels are given in parenthesis. The plots are presented on a $\log_{10}$ abscissa.

distribution over almost the entire spectrum of analytes in clinical chemistry, Figure 2A and B shows ratio histograms based on the data of Rili-BAEK 2003 (column 5/6) [7] and the database of desirable limits by Ricos et al. (version 2014) [6, 8]. In addition, all entries by Ricos, which base on more than one reference source, are again analyzed separately, due to the assumption that these entries represent well-studied techniques to a greater extent. Biases and imprecisions by Ricos et al. are derived from biological variations of each analyte (see [9, 10]). Thus, they usually

represent higher limits compared to the state-of-the-art limits of uncertainty. Particularly the limits for maximum bias are very liberal, due to the additional consideration of the entire inter-individual variation.

It can be concluded that the variation of bias-imprecision ratios is rather limited, whereas bias is still allowed to be the dominant permissible uncertainty in most cases. One might further conclude that a trend toward a less permissible bias exists in more recent data as well as for limits based on state-of-the-art (i.e. Rili-BAEK). However, the proposed optimum ratio of $\delta_{max}/s_{max} = 0.71$ [11] is still a demanding goal for almost all analytes. Anyway, the frequency histograms in Figure 2 allow two important conclusions: Due to a distinct maximum permissible bias similar or higher than s_{max}, the factor κ in eq. (4) can be chosen according to a one-sided confidence level. A value between 1.65 (95% confidence) and 2.33 (99%) is convenient. Second, the vast majority of any ratio distribution is narrow enough to estimate a general λ value for almost every analyte in clinical chemistry:

$$\lambda = 2.0 \pm 0.3. \qquad (5)$$

With regard to the entire spectrum of analytes in Ricos et al. and a large decision flexibility in κ (1.65–2.33), λ would have a maximum uncertainty of 2.00 ± 0.45. This extended uncertainty, mainly originating from differences in biological variation [9, 10], would be too liberal for limits based on state-of-the-art.

To re-establish separate limits for SMC and RA using λ, it must be clarified if the present common limit, actually declared in Rili-BAEK, can be allocated to L_{SMC} or L_{RMSTD}. On the mathematical point of view, the common limit should match to L_{RMSTD} (see [5]). However, the underlying limits of bias and imprecision (column 5/6 [7]), which were originally applied to generate the common limits in 2003, appear much too tolerant from today's perspective. Furthermore, recent personal communications with a number of large German laboratories revealed seldom amounts of SMC outliers using the present common limits of the Rili-BAEK. The current RA (RMSTD/QMMA) almost never generates an out-of-control alert in daily routine. Hence, it can be concluded that the Rili-BAEK common limit is instead close to a state-of-the-art limit for single control measures L_{SMC}. It should be mentioned that a direct derivation of an adequate L_{SMC} from the related allowable total error (TEa) given, e.g. in [6], is not straightforward. Indeed, it has been shown that limits derived from biological variation are often significantly more tolerant than necessary [12].

The situation regarding self-developed internal laboratory deviation limits is very similar. The common use of a factor 3 of the evaluated imprecision (see Rili-BAEK, part B1, chapter 2.1.4) results in too tolerant deviation limits

for RA. For SMC, on the other hand, the limits provide rather small room for tolerated increased variation.

In the next chapter, the factor λ can now be utilized to develop a general and beneficial procedure for the entire internal quality assurance with an adaptive limit.

The Adaptive Retrospective Monitoring

The novel Adaptive Retrospective Monitoring (ARM) combines all aspects of an internal quality assurance in a simple way. Utilizing the factor λ, a single formula can be applied to control single SMC results as well as retrospective on-the-fly statistics of recent SMC data. The limits used for retrospective data interpretations converge toward the stricter L_{RMSTD} with increasing amounts of considered data. Hence, a violation of the statistical distribution of recent SMC data will be recognized as fast as possible. It can also quickly respond to a suspicious sequence of SMC values, each closely within the $\pm L_{SMC}$ limits.

First, either the limit L_{SMC} or L_{RMSTD} needs to be predefined. As discussed in the previous chapter, the already established Rili-BAEK limit Δ_{max} is close to a feasible value of L_{SMC}. The extended monitoring of SMC data by ARM allows a definition of a rather tolerant limit for SMC

$$L_{SMC} = \Delta_{max} + C_\Delta = \frac{y_0}{100\%} \cdot \Delta_{max}^{rel} \cdot (1+c).\qquad(6)$$

In concert with eq. (1), y_0 is the target value of the control sample and Δ^{rel} indicates the relative limit given in Rili-BAEK, table B1a–c, column 3. The additional general constant C_Δ grants a small bonus to the present common limit to ensure a sufficiently rare appearance of outliers under error-free working conditions. It is suggested to specify an overall constant c as a small offset factor to any Δ_{max}. A proper value of c depends on the approach or criteria originally used to determine the present common limits. It is supposed to lie between 0 and 0.2. An offset above 0 would also lead to compensate the analyte-specific variation of λ around the general mean value [see eqs. (5) and (7)].

The formula of ARM (utilizing the sole limit L_{SMC}) is an extension of eq. (2) by a transition function a(n) used to adjust a proper limit value to a given sample size n.

$$\sqrt{\frac{1}{n} \cdot \sum_{i=1}^{n}(y_i - y_0)^2} \leq \frac{a(n)}{\lambda} \cdot L_{SMC} \quad \text{with}$$
$$a(n) = 1 + (\lambda - 1) \cdot \exp(-0.3(n-1))\qquad(7)$$

Thus, limits become increasingly less tolerant for larger n. A proper value for λ is given in eq. (5). On each application, in addition to the most recent SMC measure y_i, the procedure also considers n – 1 previous SMC results of the same control sample. The recognized number n of recent data is either limited by the application time of the control sample or to max. 15 most recent SMC values. After each SMC measure the ARM formula is applied n times for $\{y_i\}$, $\{y_i, y_{i-1}\}$, $\{y_i, y_{i-1}, y_{i-2}\}$, …, $\{y_i, y_{i-1}, …, y_{i-n-1}\}$ (i.e. retrospectively accumulated SMC values), leading to n potential violations of the appropriate limit level according to eq. (7). The final evaluation of all n limit tests is done by the definition of alert levels given in Table 1. Please note that a moderately off-limit SMC value will usually only be recognized by the n = 1 limit. A strong exceeding at n = 1, on the other hand, will very likely also lead to exceed further limits for n > 1. Thus, more nuanced results are possible, as shown in Table 1.

For each new SMC value, it might be sufficient to apply eq. (7) only at odd positions (i.e. n = 1, 3, 5, …, 13, 15) to retrospectively control recent SMC data. Thus, only eight comparisons to the dedicated limit value are applied considering the last 15 SMC values.

How to deal with outliers? In principle, outliers are important state indicators. They need to be additionally recognized in future retrospective statistics, even if the SMC measure was directly repeated. Anyway, if an outlier has clearly been dedicated to an apparent momentary mistake or a compromised control sample, it should be rejected. The frequency of outliers can often be reduced by shortening of the calibration period, a reduction of the measuring time of the control sample and using optimally stored aliquots of the control sample.

The concept of a static evaluation period of 1, 2 or 3 months is now obsolete. Up to 15 recent SMC values are considered permanently.

Table 1: Alert levels of the Adaptive Retrospective Monitoring.

Level	Indication	Description
0	In control	No violation of limits
1	Warning	Up to two violations except at n = 1
2	Suspicious	Outlier at n = 1 and no other violations
3	Problematic	Outlier at n = 1 and one further violation
4	Statistically out-of-control	More than two violations but no outlier at n = 1
5	Fully out-of-control	More than two violations including an outlier at n = 1
6	Deprecate lot/ device	Three consecutive level-5 events; reportable incident

Theoretical reasoning led to the conclusion that the factor 0.3 within the exponential term of a(n) ensures a sufficiently flat transition curvature, which optimally avoids false off-limit alerts in the transition range $1 < n < 15$ (nevertheless preventing too tolerant behavior). However, although not recommended, the factor can in principle be reduced up to 0.25. This factor and the best definition of alert levels, given in Table 1, still need to be validated in an extended field study. The newly developed ARM provides following features: simple calculation, effective retrospective monitoring, only one mandatory limit is still sufficient, entire internal quality control by one approach allowing more detailed interpretations of results.

Supplement to the external quality assurance

The current Rili-BAEK demands one valid certificate for successful participation in an EQA per analyte and for an entire location (unique postal address). Large laboratories with several active devices of same type, alternative detection platforms for the same analyte, or further standby devices can actually freely decide on which device EQA samples are analyzed. In a worst case scenario, always the same most trusted device (with the lowest liability to fail) is applied to measure EQA samples. Hence, the present EQA strategy cannot be considered as a quality monitoring of single devices nor for all used detection techniques. It is a pure proficiency test for the general ability to correctly measure an analyte. However, a monitoring of each device by EQA is not realistic, due to a huge demand on control material, e.g. taken from (positive) blood donors. Thus, it is suggested to keep the present regulation unchanged, although extended by a rule, claiming for a rotatory use of all active devices. Further, if an EQA participation was unsuccessful, the next EQA has to be analyzed with the same device that failed before. It might also be suggested that sufficiently stable EQA samples should be optimally stored to aid in a rekindling process of backup devices or repaired/maintained devices. Such a subsequent measure can be compared with the official results presented in the related Youden plot on own responsibility.

Group-specific and continuous reference intervals

Great effort has been spent in clinical chemistry to ensure the quality of analytic results. However, the medical usefulness of a measured result distinctly depends on the correct match of the applied reference interval for the individual patient. Even the definition of reference intervals is nontrivial. A reference interval covers (with a confidence level of 95%) the inter-individual variation of results within a representative subpopulation of healthy persons, to which the patient clearly belongs. Intended or not, some intra-individual variability (e.g. circadian rhythms, prandial state) could be left out, if all persons were examined under the same conditions (e.g. daytime, fasting/soberness, fitness). With regard to pre-analytic requirements, communication deficits between medical practitioners and the laboratory still occur and should be addressed in the Rili-BAEK more prominently.

The differentiation of entire reference groups of healthy persons into subpopulations is essential for a number of analytes, because group-specific characteristics like age, gender, genotype/ethnicity, pregnancy, lifestyle (smoking, weight/fitness, etc.), medication, etc. often influence the position and range of reference intervals distinctly. Moreover, children of a specific age could stay in rather different physical development stages. This huge amount of potential influencing factors illustrates that the required minimum amount of human reference samples to reveal separate group-specific reference intervals has often been significantly underestimated. As a consequence, group-specific intervals from different literature sources diverge or show suspect shifts in position and width of reference intervals of consecutive age groups within the same study. In this context, the explicit advice by the Rili-BAEK to generate own laboratory-specific reference intervals appears problematic [13]. Several laboratories are not able to guarantee important requirements to generate reliable intervals, due to insufficient variability in the available pool of patients, the possibility of missing related information about the patients and the blood-sampling conditions as well as the inherent problem to distinguish sick from healthy persons in a pool of patients by statistical methods. Concerning to the last point, the following has to be mentioned: Group-specific reference intervals are always superpositions of distributions of even more specific subpopulations (e.g. different genotypes or lifestyle). Thus, it cannot be assumed that the shape of a sample distribution of purely healthy individuals is smooth and without any shoulders or even local peaks. Hence, a disturbance in the distribution curve of patient samples cannot be definitely matched to the fraction of sick people.

In recent years, large international campaigns have been launched to determine group-specific reference intervals based on big data of putatively healthy

persons. Special regard goes to CALIPER (Canada) [14, 15], NHANES (USA) [16, 17], CHILDx (USA) [18], AACB (Australian) [13, 19], NORIP (Scandinavian) [20], NUMBER (The Netherlands) [21], the German projects KiGGS [22, 23], LIFE [24], PEDREF [25] and the EuBIVAS database of EFLM (Europe) [8]. Hence, several large data sets are accessible to define reliable and most comprehensive group-specific reference intervals. If method- and platform-specific variations between different data sources are identified and mostly eliminated, all available data could be combined to build a central German database of reference intervals, which are representative, reliable and standardized. Significant differences of reference intervals in local populations of Germans are rare and most probably already known on site.

If such a central database exists, manufacturers of clinical devices and reagent kits would also get a standardized and reliable reference source to minimize the platform-/lot-specific permanent bias (i.e. constant inaccuracy) of measuring systems by a more sophisticated calibration (see also below). Alternative sources to detect and quantify such biases are Youden plots of EQA programs.

Recently, international attempts has been initiated to harmonize reference intervals and cross-platform analytic results [21, 26–30]. Although the harmonization effort proved to be a complex task, the principle of metrological traceability is still an undisputed fact (despite a vague critical comment in [31]). However, the current approach to maintain metrological traceability (via a cascade of calibrators of decreasing ranks) might be finally unsuited to prevent distinct platform-/lot-specific biases, probably due to accumulating measurement errors or different matrix characteristics of calibrator samples. Such inherent problems lead to the need of advanced calibration or re-calibration methods. This subjective statement is based on the conviction that results in laboratory medicine should be primarily harmonized (which means consistent compared to alternative approaches as well as the related public reference interval) rather than putatively more "true" than the result of an alternative measuring approach. To the knowledge of the author, control material with a target value denoted as "Sollwert" (nominal value) can actually be excluded to assess full traceability and trueness, if the control material is only intended to verify the measuring consistency over time and the reliability of the measuring procedure. This regulation may hide the real inaccuracy of calibrators at the final end of the metrological "cascade". This could be evaluated by comparative measures of control material dedicated to another lot or platform. It must further be noted that reported reference intervals not always reflect the full amount of the real platform-/lot-specific bias of the actual result.

Harmonized and reliable group-specific reference intervals with continuous transitions between consecutive age groups are also essential prerequisites for the "individualized medicine" and the implementation of unitless results in clinical chemistry like the zlog value [32].

In near future, the amount of individual laboratory data of patients will further increase and will also be generally available via an electronic card or on a central cloud. Hence, the within-subject longitudinal monitoring of analytes becomes increasingly practicable and important. Comparing intra- and inter-individual biological variations, it is well known that several analytes show a very narrow within-subject deviation compared to their between-subject variation (i.e. the entire reference interval). Prominent examples with an intra- vs. inter-individual ratio of biological variation of less than 0.25 are α-fetoprotein, α2-macroglobulin, alkaline phosphatase, C and S protein, collagen propeptides, creatine kinase (mass), dehydroepiandrosterone sulfate (DHEAS), thyroid antibodies, troponin I, vitamin B12, von Willebrand factor, some amino acids, most of the cancer antigens, etc. [6]. Moreover, the relevant C3/C4 complements are slightly above (<0.33). Thus, it is important to consider that a patient result could reveal an alarming analyte deviation with respect to the historic personal healthy range of the patient, although the result is still within the general reference interval [33]. It is therefore recommended to additionally provide a table of stricter decision limits especially dedicated to longitudinal monitoring of important analytes with very low ratios of intra- vs. inter-individual variation. The mathematical basis for such limits is the known reference change value (RCV) [34]. Here again, distinct lot-/platform-specific biases unnecessarily complicate comparisons of recorded values from different providers.

References

1. Neufassung der "Richtlinie der Bundesärztekammer zur Qualitätssicherung labormedizinischer Untersuchungen – Rili-BÄK". Dt Aerzteblatt 2014;111:A1583–618.
2. Revision of the "Guideline of the German Medical Association on Quality Assurance in Medical Laboratory Examinations – Rili-BAEK" (unauthorized translation). J Lab Med 2015;39:26–69.
3. Richtlinie der Bundesärztekammer zur Qualitätssicherung quantitativer labormedizinischer Untersuchungen. Dt Aerzteblatt 2001;98:A2747.
4. Richtlinie der Bundesärztekammer zur Qualitätssicherung labormedizinischer Untersuchungen. Dt Aerzteblatt 2008;105:A341–55.
5. Mcdonald R. Quality assessment of quantitative analytical results in laboratory medicine by root mean square of measurement deviation. J Lab Med 2006;30:111–7.
6. Desirable Biological Variation Database specifications. https://www.westgard.com/biodatabase1.htm. Accessed: 15 Jan 2019.
7. Richtlinie der Bundesärztekammer zur Qualitätssicherung labormedizinischer Untersuchungen. Dt Aerzteblatt 2003;100:A3335–8.
8. EFLM. https://biologicalvariation.eu/bv_specifications/measurand. Accessed: 26 May 2019.
9. Biswas SS, Bindra M, Jain V, Gokhale P. Evaluation of imprecision, bias and total error of clinical chemistry analysers. Ind J Clin Biochem 2015;30:104–8.
10. Fraser CG. Biological variation: from principles to practice. Washington, DC: AACC Press, 2001:1–151.
11. Haeckel R, Gurr E, Hoff T. Bias, its minimization or circumvention to simplify internal quality assurance. J Lab Med 2016;40:263–70.
12. Oosterhuis WP. Gross overestimation of total allowable error based on biological variation. Clin Chem 2011;57:1334–6.
13. Graham P, McLemon E. Harmonisation of adult and paediatric reference intervals for general chemistry analytes across Australian pathology laboratories. 2014. Report of the Harmonised Reference Intervals Project.
14. CALIPER. http://www.sickkids.ca/Caliperproject/index.html. Accessed: 26 May 2019.
15. Higgins V, Adeli K. CALIPER database of paediatric reference intervals: key milestones and future direction. Folia Medica Copernicana 2005;3:7–12.
16. NHANES. https://www.cdc.gov/nchs/nhanes/index.htm. Accessed: 26 May 2019.
17. Lim E, Miyamura J, Chen JJ. Racial/ethnic-specific reference intervals for common laboratory tests: a comparison among Asians, Blacks, Hispanics, and White. Hawai'i J Med Pub Health 2015;74:302–10.
18. Wyness SP, Roberts WL, Straseki JA. Pediatric reference intervals for four serum bone markers using two automated immunoassays. Clin Chim Acta 2013;415:169–72.
19. Australasian Harmonisation Project. https://www.aacb.asn.au/resources/laboratory-test-databases. Accessed: 26 May 2019.
20. Rustad P, Felding P, Franzson L, Kairisto V, Lahti A, Mårtensson A, et al. The Nordic Reference Interval Project 2000: recommended reference intervals for common biochemical properties. Scand J Clin Lab Invest 2004;64:271–84.
21. den Elzen WP, Brouwer N, Thelen MH, Le Cessie S, Haagen I-A, Cobbaert CM. NUMBER: standardized reference intervals in the Netherlands using a "big data" approach. Clin Chem Lab Med 2019;57:42–56.
22. KiGGS. https://www.kiggs-studie.de/deutsch/home.html. Accessed: 26 May 2019.
23. Kohse KP, Thamm M. KiGGS – the German survey on children's health as data base for reference intervals. Clin Biochem 2011;44:479.
24. LIFE. http://life.uni-leipzig.de/de/life_forschungszentrum/life_datenportal_ldp.html. Accessed: 26 May 2019.
25. PEDREF. https://www.pedref.org/reference-intervals.html. Accessed: 26 May 2019.
26. Myers GL, Miller WG. The roadmap for harmonization: status of the International Consortium for Harmonization of Clinical Laboratory Results. Clin Chem Lab Med 2018;56:1667–72.
27. Miller WG, Myers GL, Gantzer ML, Kahn SE, Schönbrunner ER, Thienpont LM, et al. Roadmap for harmonization of clinical laboratory measurement procedures. Clin Chem 2011;57:1108–17.
28. Tate JR, Koerbin G, Adeli K. Opinion paper: deriving harmonized reference intervals – global activities. eJIFCC 2016;27:48–65.
29. Koerbin G, Sikaris K, Jones GR, Flatman R, Tate JR. An update report on the harmonization of adult reference intervals in Australasia. Clin Chem Lab Med 2018;57:38–41.
30. Plebani M, Graziani MS, Tate JR. Harmonization in laboratory medicine: blowin' in the wind [Editorial]. Clin Chem Lab Med 2018;56:1559–62.
31. Oosterhuis WP, Bayat H, Armbruster D, Coskun A, Freeman KP, Kallner A, et al. The use of error and uncertainty methods in the medical laboratory. Clin Chem Lab Med 2018;56:209–19.
32. Hoffmann G, Klawonn F, Lichtinghagen R, Orth M. Der zlog-Wert als Basis für die Standardisierung von Laborwerten. J Lab Med 2017;41:23–32.
33. Walz B, Fierz W. The concept of reference change values (RCV). Will it supersede reference intervals? Ther Umsch 2015;72:130–5.
34. Ricos C, Cava F, Garcia-Lario JV, Hernandez A, Iglesias N, Jimenez CV, et al. The reference change value: a proposal to interpret laboratory reports in serial testing based on biological variation. Scand J Clin Lab Invest 2004;64:175–84.