Klin. Biochem. Metab., 25 (46), 2017, No. 4, p. 167–176.

The S model: Method performance specifications based on Six Sigma metrics

Huba P., Vaňuga A., Dančová K.

Alpha Medical s.r.o, Department of Clinical Chemistry and Haematology, Lubochna, Slovakia

SUMMARY

Objective: Proposal of the analytical method performance requirements definition based on the Six Sigma metric.

Settings: Alpha Medical s.r.o, Department of Clinical Chemistry and Haematology, Skolska 153/2, 034 91 Lubochna, Slovakia

Material and Methods: The proposed model is based on the calculation and transformation of the Z-score. The allowable error margin was determined as a difference between Z=3.145 and Z=3, according to the Six Sigma metric. All calculations were performed in Microsoft Excel 2010.

Results: We found that any combination of the square root of the sum of squares of CV_A and B_A related to a biological signal equal to 0.379, is leading to the maximum bi-directional error rate of 6.68 %.

Conclusion: The maximum allowable analytical error, resulting from the S model, causes a maximum additional increase of 1.68 % of patient results outside a reference interval. Since the proposed model is two-sided, it allows defining both maximum mutually exclusive criteria for imprecision and bias and their other possible combinations. The S model uses a simple quadratic formula with one cut off value. Besides the so-called reference approach solely based on the biological variation, we propose also the state-of-the-art approach of defining method performance specification.

Keywords: quality specifications, performance goals, Sigma metrics, total allowable error, biological variation.

SOUHRN

Huba P., Vaňuga A., Dančová K.: S-model: Návrh určení výkonnostních požadavků na analytické metody založeného na Six Sigma metrikách

Cieľ štúdie: návrh stanovenia výkonnostných požiadaviek na analytické metódy založeného na Six Sigma metrike

Názov a sídlo pracoviska: Alpha medical s.r.o., Laboratórium klinickej biochémie a hematológie, Školská 153/2, 034 91 Ľubochňa, Slovensko

Materiály a metódy: Navrhovaný model je založený na výpočte a transformácii Z skóre. Najväčšia prípustná chyba merania bola určená ako rozdiel medzi Z=3,145 a Z=3 na základe Six Sigma metriky. Všetky výpočty k práci boli uskutočnené v prostredí Microsoft Excel 2010.

Výsledky: Z našej práce vyplynulo, že akákoľvek kombinácia bias a precíznosti merania v pomere k biologickému signálu, v tvare ich druhej odmocniny zo sumy štvorcov, rovná výsledku 0,379, vedie k maximálnemu, obojstrannému, 6,68 % podielu pacientskych výsledkov mimo referenčný interval.

Záver: Maximálna povolená celková analytická chyba, vyplývajúca z S modelu, spôsobuje navýšenie maximálne o 1,68 % pacientskych výsledkov mimo referenčný interval oproti očakávaným 5 %. Vzhľadom na to, že navrhovaný model berie do úvahy obe strany referenčného intervalu, dovoľuje definíciu jednak vzájomne sa vylučujúcich podmienok pre precíznosť a bias ako aj ich iné, možné kombinácie. Navrhovaný model je založený na jednoduchej kvadratickej rovnici s jednou cut off hodnotou. Okrem postupu založeného čisto len na údajoch o biologickej variácii, ktorý je považovaný za referenčný, v práci navrhujeme aj tzv. state-of-the-art postup.

Kľúčové slová: požiadavky na kvalitu, výkonnostné ciele, sigma metrika, celková povolená chyba, biologická variácia, výkonnostné ciele.

Introduction

The quality of laboratory test results is crucial for physicians to make decisions on a person's health status. Setting appropriate analytical quality specifications in clinical laboratories is a demanding task. The most common way to define these specifications is based on biological variation (BV) which is recognized by the scientific community as the second approach in the three-stage hierarchy resulting from the 1st Strategic Conference of the European Federation of Clinical Chemistry and Laboratory Medicine held in Milan in 2014 [1].

Different approaches have been discussed how to derive quality specifications from BV [2-6] and commonly accepted model is based on expressions (1), (2).

$$CV_A \leq 0.5 \times CV_I \qquad (1)$$

$$B_A \leq 0.25 \times \sqrt{(CV_I^2 + CV_G^2)} = 0.25 \times CV_B \qquad (2)$$

Where,

B_A= bias

CV_A = analytical variation (imprecision)

CV_I = within-subject biological variation

CV_G = between subject biological variation

CV_B = total biological variation

These conditions, are combined with an expression of total error allowable (TE_A), (3), [7].

$$TE_A = Z \times CV_A + B_A \qquad (3)$$

Where,

Z is the coverage factor and generally is taken as 1.65

for a one-sided estimate for a 95 % confidence interval.

The concept of total error in clinical chemistry first introduced Wetsgard in the 1970s. A complete list of biological variation with corresponding performance specifications is published and updated on the Westgard website [8]. These specifications, particularly TE_A, are largely used in daily practice by clinical laboratories to set quality control limits. The concept of total error is subjected to criticism, and pros and cons of the TE_A need to be debated [6, 9, 10]. Here we list some of the flaws of the concept.

The conventional model assumes a linear relationship of analytical error components. This assumption is valid only if a biological signal doesn't play a role. In fact, the relationship between these components related to BV is nonlinear and can be demonstrated by a curve (Figure 1).

Conditions based on expressions (1), (2) are not related to the same subject. In the case of CV_A, the performance goal is derived from within-subject biological variation and relates to patient monitoring. In the case of bias (B_A), the performance goal is derived from total BV and relates to diagnosis. The combination of these different performance goals makes this model the CV_I/CV_G ratio (index of individuality) dependent (see also Appendix 3).

The performance goals defined by the expressions (1), (2) are mutually exclusive and are valid only under the assumption that CV_A and B_A equal to zero, respectively. Thus, this model doesn't provide an information what other combination of imprecision and bias is acceptable.

Let's take a look at an example, e.g., of serum sodium. From a biological variation database [8] we can obtain these data: CV_I = 0.60 %, CV_G = 0.70 %, CV_B = 0.92 %.

According to the expressions (1), (2), and (3), the desirable specifications will be as follows: CV_A = 0.30 %, B_A = 0.23 %, TE_A = 0.73 %

Now, consider that in a clinical laboratory an imprecision ($CV_{A, LAB}$=0.50 %) and bias ($B_{A, LAB}$=0.10 %) were determined. In the view of the conventional model, imprecision and total error allowable are unsatisfactory since the predefined limits were not met. It should be noticed again that performance goals defined by the expressions (1), (2) are mutually exclusive and are valid only under the assumption that CV_A and B_A equal to zero, respectively. According to this, the addition of TE_A is a little of value. Further, what could we say about observed $B_{A, LAB}$ =0.10 %? In the case the imprecision would be close to zero, the bias criterion would be met. What imprecision, with observed $B_{A, LAB}$ = 0.10 %, is still acceptable? With respect to the expression (3) we can try to compute allowable CV_A as follows:

$$CV_A = \frac{(TE_A - B_{A, LAB})}{1.65} = \frac{(0.73 - 0.10)}{1.65} = 0.38\%$$

From this example, it is clear that computed allowable imprecision of 0.38 % is higher than the predefined criterion of 0.30 %. Thus, the expression (3) doesn't provide reliable information what imprecision would be acceptable with observed bias and vice versa. Note that, this is a hypothetical example. The observed bias and imprecision stated above are made up and don't reflect the reality. In fact, the assumptions resulting from the conventional model for sodium are so demanding, that cannot be fulfilled with contemporary technology.

The most important question about any model or any performance specification is what effect will, an analytical error on patient results have if the specifications would be met. Since the conventional model doesn't provide an answer to this question, we would like to introduce a different approach to define the analytical method performance requirements with a direct link to patient results.

The reference interval (RI) concept

Comparison of a laboratory test against a reference is the main criteria for making medical decisions. Dependent on a laboratory tests nature, this reference could be a population RI (one or two sided), previous test result or in some cases the decision limits set by national or international consensus (glucose, cholesterol, CRP, HbA1c etc.) are used. The concept of RI is a broader issue which is beyond the scope of this paper. Here we discuss only the essence of RI needed for the understanding of the proposed model.

The RI is the interval specified in the distribution of values obtained from populations of healthy subjects. This is generally defined as an interval corresponding to 95 % of the population centered on the median [11]. In the case of normally distributed data, RI corresponds to ±1.96 SD around the mean. From the definition of RI, it is clear, that bi-directional error rate (ER_B) or patients outside the RI equals to 5 %. The width of a population-based reference interval in term of a coefficient of variation (CV_{RI}), is defined mostly by biological variation and analytical variation of a method used to measure reference values:

$$CV_{RI} = \sqrt{(CV_I^2 + CV_G^2 + CV_{A0}^2)} \qquad (4)$$

Where CV_{A0} is the analytical variation of a method when the reference limits were determined. Since the biological variation is inherent, the maximum effort should be spent to minimize the contribution of analytical variation. In an ideal situation, the analytical variation would be neglected and the width of a population RI becomes defined only by biological variation itself:

$$CV_{RI} = CV_B \leq \sqrt{(CV_I^2 + CV_G^2)} \qquad (5)$$

The usefulness of a population RI is not straightforward. The main disadvantage is that a low variation in test result may be of pathophysiological importance even if the result is within the reference range. To overcome this issue, the best option would be the use of a personal RI.

$$CV_{RI} = CV_I \qquad (6)$$

Obtaining the personal RI for every individual is apparently impossible. Rather than personal RI, the concept of reference change value (RCV) is used for monitoring significant changes in laboratory results from an individual [12].

$$RCV = \sqrt{2} \times Z \times \sqrt{(CV_I^2 + CV_A^2)} \qquad (7)$$

The Six Sigma methodology

The Six Sigma methodology was introduced by a company of Motorola as a tool for lowering process variability, a price of a products and defect rate.

When we think about Six Sigma as a metric, the term sigma is used as a scale for level of quality and can range from -∞ (infinity) to +∞. The higher the sigma level, the better the process is performing and the lower the probability that a defect will occur. Statistically, the Six Sigma refers to a process in which the range between the mean of a process quality measurement and the nearest specification limit is at least six times the standard deviation of the process. Sigma quality level is usually indicated by the letter Z. One of the common measures of a process's performance in Six Sigma is Defects per Million Opportunities (DPMO). A Six Sigma process will approach "zero defects" with only 3.4 of DPMO for a defect to occur. It should be noted that Six Sigma differentiates between short- and long-term process variation. A value of Z= 6 in a normal distribution table really translates to 2 defects per billion opportunities (two sided) and a DPMO of 3.4 originally corresponds to a Z=4.5 (one sided). The corresponding DMPO's differ significantly and this is due to that long-term sigma metric is shifted against short-term by 1.5 sigma. This difference is advocated by means of long-term process variation. Motorola empirically determined that processes vary over a long period of time and an average of this variability equals to 1.5 sigma [13]. Thus, short-term sigma performance is higher about 1.5 sigma then long-term sigma ($Z_{ST}=Z_{LT}+1.5$). In a Six Sigma conversion table (Table 1), the sigma level of performance is defined as short- term sigma along with projected performance behavior in term of DPMO offset by 1.5 sigma (long-term DPMO, one-sided).

The Sigma model (The S model)

The S model is a two-sided approach for a definition of analytical performance goals. It should be noted that this model assumes a Gaussian distribution only. When imprecision increases (assuming no bias is observed) the number of patient results outside the RI will symmetrically increase on both sides of RI. A shift of a method's performance will cause asymmetrically distribution of results outside the RI. The S model is constructed in the way that in every situation, given in maximum CV_A, B_A or in any combination, the contribution of an analytical error leads to the same portion of patient results outside RI as a sum of the portion of the both sides of a distribution. To determine the minimum allowable specification, the Sigma metrics approach was used as follows.

Initially, as a minimum acceptable level, 3 sigma values (Z=3) were taken into account. Generally, this level of sigma is considered as the minimum performance of processes [13]. Expressed as DPMO, this level of sigma performance relates approximately to 66800 (6.68 %). From the definition of RI (95 % central interval) it is clear that 5 % of patient results will lie outside RI (2.5 %/ 2.5 % below and above limit, respectively), (Figure 2, case A). Expressed as DPMO it means 50000 results. A DPMO of 50000, in the Six Sigma conversion table, relates to a value of Z= 3.145. In the sense of Six Sigma metrics, we can say that RI operates at the level of 3.145 sigma.

According to thoughts above, the amount of the acceptable error can be determined as a difference in 66800 DPMO (Z=3.145) and 50000 DPMO (Z=3) which equals to DPMO of 16800. Let's call this an error margin (EM). In other words, the S model considers the contribution of an analytical error that causes a maximum additional increase of 1.68 % of patient results outside the original distribution. In such a situation bi-directional error rate, ER_B= 6.68 % will be observed against expected 5 %. Every combination of an analytical error component in term of CV_A/CV_B and B_A/CV_B that fulfill this assumption is considered to be acceptable. A maximum bias (8) and imprecision (9) in

Table 1: Six Sigma conversion table (rounded to three significant digits)

Sigma scale (Z_{ST})	Short term DPMO (centered distribution, two-sided)	Long term DMPO (shifted distribution = offset of 1.5 SD, one-sided)
0	**1000000**	SHIFT
0.5	617000	
1	31700	
1.5	134000	**500000**
2	45500	30900
3	2700	66800
4	63.3	6210
5	0.574	233
6	0.002	3.40

term of a ratio to BV was found to fulfill the assumption of ER_B= 6.66 and 6.68 %, respectively (see also Appendix 1).

A maximum allowable bias with zero imprecision can be expressed as:

$$\frac{B_A}{CV_B} = 0.379 \qquad (8)$$

ER_B =6.66 % (0.97 %/ 5.69 % below and above limit, respectively), (Figure 2, case B)

A maximum allowable imprecision with zero bias can be expressed as:

$$\frac{CV_A}{CV_B} = 0.379 \qquad (9)$$

ER_B =6.68 % (3.34 %/ 3.34 % below and above limit, respectively), (Figure 2, case C)

Let's call the coefficient B_A/CV_B, CV_A/CV_B as an error margin coefficient (EM_C).

For many other possible combinations, that maintain the predefined limit, the Pythagorean Theorem (10) can be used to check whether observed bias and imprecision in clinical laboratory fulfill the minimum specification according to the S model (Fig. 1).

$$c^2 = a^2 + b^2 \qquad (10)$$

The general expression for the S model can be expressed as:

$$EM_C \le 0.379 \le \sqrt{\left(\frac{CV_A}{CV_B}\right)^2 + \left(\frac{B_A}{CV_B}\right)^2} \qquad (11)$$

For bias and imprecision synergistic effect illustration, consider this allocation of the analytical error:

$$\frac{B_A}{CV_B} = 0.286; \qquad \frac{CV_A}{CV_B} = 0.249$$

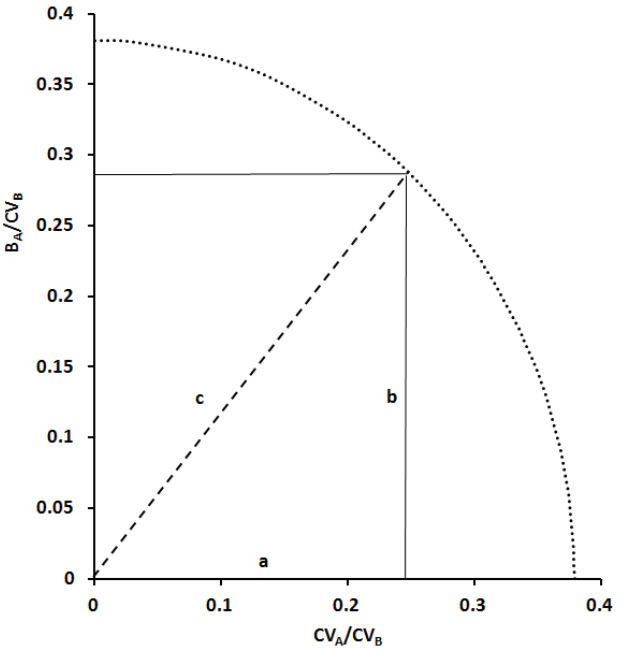

Figure 1. The curved relation between bias and imprecision related to BV with the maximum bi-directional error rate of 6.68% according to the S model.

$$EM_C = \sqrt{\left(\frac{CV_A}{CV_B}\right)^2 + \left(\frac{B_A}{CV_B}\right)^2}$$

$$EM_C = \sqrt{0.286^2 + 0.249^2} = 0.379$$

ER_B = 6.69 % (1.47 %/ 5.22 % below and above limit, respectively), (Fig. 2, case D).

Up to here, the width of RI was under consideration that is determined by total BV and thus relates to diagnosis. In the case of monitoring the same principle can be used with the exception that, in the expression (11), CV_B is replaced by CV_I. This approach, based only on biological variation, can be considered as the "reference approach".

Table 2 summarizes the effect of a maximum allowable imprecision and bias to bi-directional and uni-directional error rate (ER_U) according to the S model).

Using the S model, in contrast to the conventional model, judging the method performance is more comprehensible. All we need to do is to compare observed characteristics (imprecision, bias), to CV_I, CV_B. In our example of serum sodium, mentioned in the introduction, the situation will be as follows:

$$CV_{A,\,LAB} = 0.50\%,\ B_{A,\,LAB} = 0.10\%$$

$$\frac{CV_{A,\,LAB}}{CV_I} = \frac{0.50}{0.60} = 0.833; \qquad \frac{B_{A,\,LAB}}{CV_I} = \frac{0.10}{0.60} = 0.167$$

$$\frac{CV_{A,\,LAB}}{CV_B} = \frac{0.50}{0.92} = 0.543; \qquad \frac{B_{A,\,LAB}}{CV_B} = \frac{0.10}{0.92} = 0.109$$

According to the expression (11), the EM_c will be:
For monitoring

$$EM_C = \sqrt{0.833^2 + 0.167^2} = 0.850$$

For diagnosis

$$EM_C = \sqrt{0.543^2 + 0.109^2} = 0.554$$

The computed $EM_{C's}$ are higher than the minimum requirement ($EM_C \le 0.379$) and therefore the method performance of serum sodium is inappropriate for neither monitoring nor diagnosis purposes. In fact, the requirements for serum sodium from BV are so demanding that they are unlikely to be fulfilled with contemporary technology. Generally, this is a well-known fact.

The state-of-the-art approach of method performance specifications

Although meeting the performance specifications based on BV (CV_I, CV_B) is preferable, in some cases, e.g., serum sodium, chloride, fT4 just to name a few, we need to go down in performance goals hierarchy and use the state-of-the-art approach. In such a situation, the main objective is the maintenance of applicability of common RI. It should be noted that this concept assumes a Gaussian distribution only. The width of a common reference interval can be expressed by the formula (4).

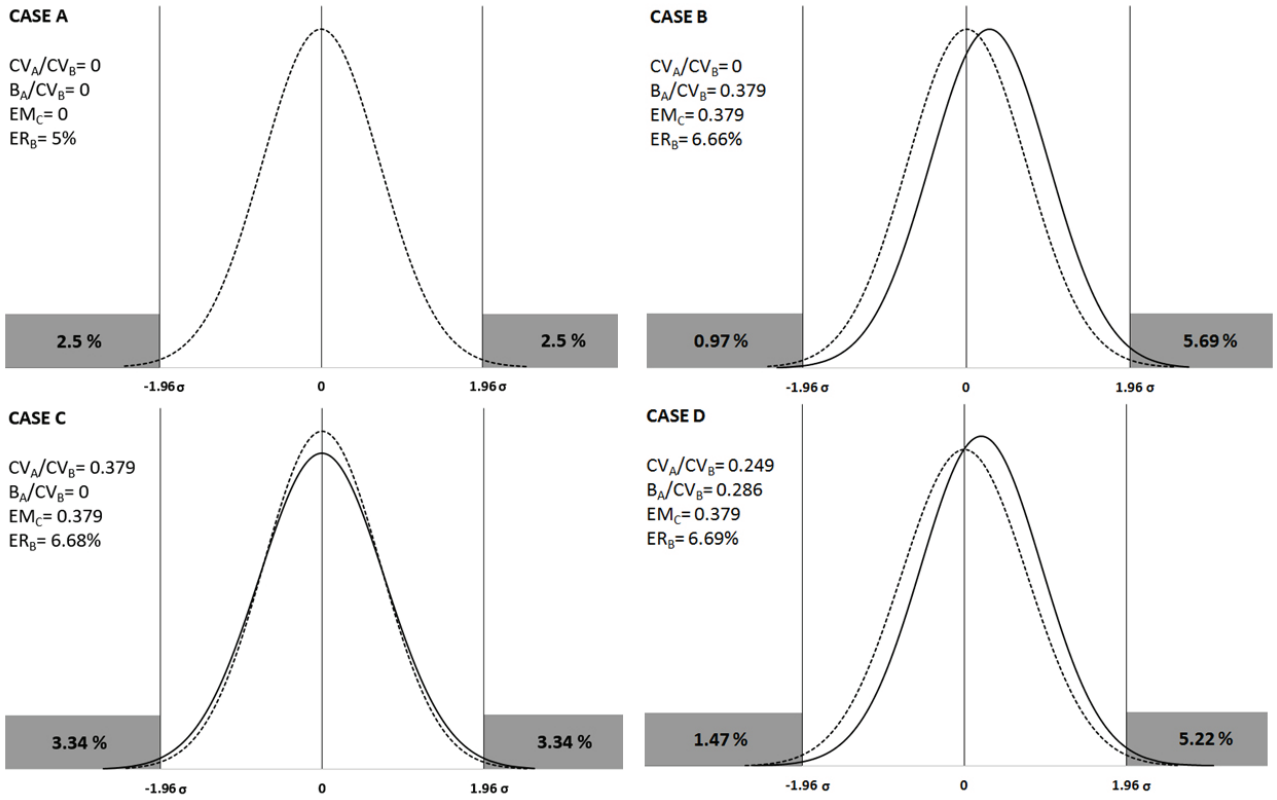

Figure 2. A Gaussian distribution with bi-directional error rates, according to the S model. The dotted curve represents a reference distribution, and the solid line curve represents an error loaded distribution. Case A- a reference distribution. Case B- the effect of maximum allowable bias, B_A/CV_B=0.379. Case C- the effect of maximum allowable imprecision CV_A/CV_B=0.379. Case D- example of synergistic effect of bias and imprecision, $EM_C = \sqrt{((B_A/CV_B)^2 + (CV_A/CV_B)^2)} = 0.379$

Table 2: Minimum acceptable performance of a method expressed as EM_C with the corresponding bi- and unidirectional error rates.

EM_C	ER_B (%) for $CV_{A,\,max}$	ER_B (%) for $B_{A,\,max}$	Average ER_B (%)	Average ER_B (%) – expected 5%	ER_U (%) for $CV_{A,\,max}$		ER_U (%) for $B_{A,\,max}$	
					Below limit	Above limit	Below limit	Above limit
0.379	6.68	6.66	6.67	1.67	3.34	3.34	0.97	5.69

$$CV_{RI} = \sqrt{CV_I^2 + CV_G^2 + CV_{A0}^2}$$

Where CV_{A0} is the analytical variation of a method when the reference limits were determined. The S model, however, cannot be used here because an allowable analytical error becomes defined by itself. We cannot use the width of common RI in that way as in the "reference approach". The simplest case is when laboratory determines its own reference interval and the CV_{A0} it is known and recorded. Then, if the bias remains neglected, the analytical requirements will be determined by CV_{A0}, that is a current imprecision should be equal or lower than CV_{A0}. When the bias becomes significant in time, the error budget=CV_{A0} needs to be allocated both into bias and imprecision to ensure that no more than 2.5 % of patient results will lie outside the common RI, one-sided. If the whole error budget applies to bias, i.e. B_A/CV_{A0}= 1, CV_A/CV_{A0}=0, the applicability of the common RI could be compromised. In fact, allocation of bias will depend on the CV_{A0}/CV_{RI} ratio. If CV_{A0}/CV_{RI} >0.81, then B_A/CV_{A0} could be >1, and vice versa (Fig. 3).

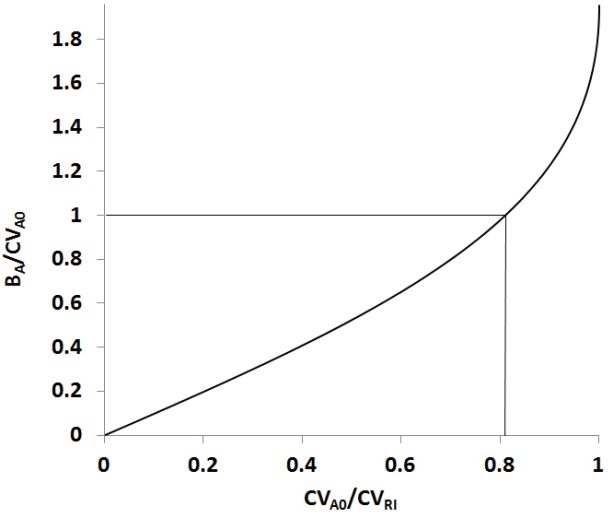

Figure 3: Relation between CV_{A0}/CV_{RI} and B_A/CV_{A0} ratio.

The maximum allowable bias can be expressed as (see also Appendix 2):

$$B_A \leq 1.96 \times (CV_{RI} - CV_T) \qquad (12)$$

Where CV_T is test results variation in a reference population. For illustration, consider this example. A laboratory, following the recommendation for the reference interval determination, established its own study on RI of an analyte, let's say serum sodium. The RI of 135 - 145 mmol/l was determined.
UL (+1.96 SD, upper limit of RI) = 145 mmol/l
LL (-1.96 SD, lower limit of RI) = 135 mmol/l
Mean= 140 mmol/l
CV_{A0} = 1.57 %, B_{A0} = 0
CV_{RI} then equals to:

$$CV_{RI} = \frac{100 \times (UL - LL)}{3.92 \times mean} = 1.82\%$$

Let's say the actual method imprecision and bias is CV_A= 0.70 %, B_A= 0.40 %, respectively. Now, we need to figure out, if this allocation of the analytical error, doesn't compromise the validity of the common RI. According to the expression (12):

$$CV_{RI} \geq \left(\frac{B_A}{1.96}\right) + CV_T$$

$$CV_G = 0.70\%, \quad CV_I = 0.60\%$$

$$CV_B \leq \sqrt{(CV_I^2 + CV_G^2)} = 0.92\%$$

$$CV_T \leq \sqrt{(CV_B^2 + CV_A^2)} = 1.20\%$$

$$CV_{RI} \geq \left(\frac{0.40}{1.96}\right) + 1.20 \geq 0.2 + 1.2 \geq 1.4\%$$

Since 1.4 %<1.82 %, the actual method's performance is satisfactory.

Discussion

In this paper, we introduced a different sight to a method performance goal specification. The conventional model has flaws that undermine his validity. The S model for setting analytical performance specifications presented here is the most similar to the model described by Gowans et al. [4]. This model specified a maximum allowable imprecision and bias in relation to patient results outside RI as the S model does.

The model of Gowans et al. is based on a unidirectional assumption that a maximum of 4.6 % of the reference values is outside the limits according to IFCC recommendations on RI. Two thresholds can be set from this approach, a maximum imprecision (with B_A=0) of 0.597 x CV_B and a maximum bias (with CV_A=0) of 0.275*CV_B. Between these extreme values, many possible combinations of CV_A and B_A exist that maintain the predefined limit of 4.6 %, but as in the case of the conventional model mentioned in the introduction of this paper, it is not straightforward to obtain these combinations without constructing a curve and interpolating the data. In the case of maximum CV_A, ER_B= 9.24 % (4.62 % /4.62 % below and above limit, respectively) is observed. In the case of a maximum B_A, ER_B = 5.87 % (1.27 % /4.6 % below and above limit, respectively, or vice versa) is observed. According to the model of Gowans et al., a maximum allowable bias or imprecision causes the same unidirectional error rate (4.6 %). But, in the sense of the bi-directional error rate, the maximum allowable CV_A leads to 1.57 time greater portion of patient results outside RI in comparison to maximum allowable bias.

In summary, both the conventional and model of Gowans et al., defines mutually exclusive conditions for imprecision and bias. Other possible combinations of imprecision and bias are possible, but these models don't provide a straightforward solution to their determination. In the sense of patient result outside RI, the conventional model provides inconsistent outcomes in both uni- and bi-directional error rates. In the case of the model of Gowans et al., only one side of a distribution matters.

The proposed S model overcomes these issues. It uses a very simple formula with one cut-off value which is easy to memorize. Using the reference approach (BV database) of the S model, it is possible to judge a method's performance between different laboratories and to set minimum performance requirements related to a biological signal. Central to this approach is the reliability of data in BV database. It is known there are some issues that undermine the value of BV database and they are being considered by the European Federation of Clinical Chemistry and Laboratory Medicine, the biological variation working group (BVWG), in collaboration with a Spanish group responsible for the database updating. Currently, BVWG is working on a new sample collection in order to update the existing BV data [14].

Following the state-of-the-art approach (common RI), individual laboratories can make decisions about the quality of their methods in relation to the maintenance of the applicability of common reference intervals.

It should be noted that S model is based on the assumption of Gaussian distributed data. For those tests with known non-Gaussian distribution (e.g. TSH) or for those where decision limits are used (cholesterol, glucose, CRP, etc.) the S model gives false results unless the CV_B, CV_{RI} are obtained from properly transformed non-Gaussian reference data or population-based RI, respectively.

Finally, the way how was the maximum allowable error determined (16800 DPMO), could be questioned. We used the Six Sigma metric, which was developed primarily for the industry and in this way it could be considered as arbitrary. Coskun et al., advice the performance of clinical laboratories should be evaluated by short-term sigma metric without 1.5 sigma shift. The authors claim that if we measure long-term variations and take these into sigma metric calculations, we do not need to include the extra 1.5 sigma values. Moreover, if laboratory equipment is frequently calibrated and

reference materials are used, the bias can be measured and corrected, and again, there is no need to use a 1.5 sigma shift [15]. Similarly, and yet another, concerns about transferring principles of Six Sigma to clinical laboratories postulated Budina in his article [16]. We cannot fully agree the 1.5 sigma shift could be omitted in overall. Reference materials exist only for limited analytes, they are expensive and we beg to leave the state the laboratories are using them minimally, if at all. Further, as Budina stated, it should be clear what does it mean long-term in clinical laboratory? Is it a month, a year, a decade, etc.? Quality control samples have limited stability and they are often not commutable. Lot-to-lot variations in calibrators, reagents, control samples and other factors create a good opportunity that a shift could occur [17, 18]. We performed a simple check and looked at our quality controls results for 38 analytes from January 2016 to June 2017. First, 20 measurements at least were used to calculate mean for each level and analyte. Then we recalculated the means after almost 1.5 years of quality controls lifespan and we found they differ by the 0.48 sigma in average. What shift would be observed during a longer period of time? Our opinion is that even making our best effort, we cannot completely avoid a shift. Finally, whether the Six Sigma methodology is controversial for use in the field of clinical chemistry or not, it doesn't undermine the validity of the proposed model. Another specification instead of 6.68 % (1.68 %), could be used without changing the fundamentals of the S model.

Conclusion

The proposed S model is a simple approach of defining method performance criteria directly linked to the percentage of patient results outside the reference interval.

List of Abbreviations

B_A	bias
BV	biological variation
BVWG	biological variation working group
CV_A	analytical variation
CV_{A0}	analytical variation of a method when the reference limits were determined
CV_B	total biological variation
CV_G	between subject biological variation
CV_I	within-subject biological variation
CV_{RI}	width of reference interval in term of coefficient variation
DPMO	defects per million opportunities
EM	error margin
EM_C	error margin coefficient
ER_B	bi-directional error rate
ER_U	uni-directional error rate
LL	lower limit of a reference interval
RCV	reference change value
RI	reference interval(s)
SD	standard deviation
SD_B	standard deviation of reference values
SD_{RI}	standard deviation of reference interval
SD_T	total standard deviation of test results in a reference population
TE_A	total error allowable
UL	upper limit of a reference interval
Z	Z-score
Z_{ST}	short- term sigma scale
Z_{LT}	long-term sigma scale

References

1. **Sandberg, S., Fraser, C. G., Horvath, A. R., et al.** Defining analytical performance specifications: consensus statement from the 1st Strategic Conference of the European Federation of Clinical Chemistry and Laboratory Medicine. *Clin. Chem. Lab. Med.*, 2015, 53, p. 833–835.
2. **Cotlove, E., Harris, E. K., Williams, G. Z.** Biological and analytic components of variation in long-term studies of serum constituents in normal subjects. 3. Physiological and medical implications. *Clin. Chem.*, 1970, 16, p. 1028–1032.
3. **Harris, E. K.** Statistical principles underlying analytic goal-setting in clinical chemistry. *Am. J. Clin. Pathol.*, 1979, 72, p. 374–382.
4. **Gowans, E. M., Hyltoft-Petersen, P., Blaabjerg, O., Horder, M.** Analytical goals for the acceptance of common reference intervals for laboratories throughout a geographical area. *Scand. J. Clin. Lab. Invest.*, 1988, 48, p.757–764.
5. **Klee, G. G.** Tolerance limits for short-term analytical bias and analytical imprecision derived from clinical assay specificity. *Clin. Chem.*, 1993, 39, p. 1514-1518.
6. **Oosterhuis, W. P., Sandberg, S.** Proposal for the modification of the conventional model for establishing performance specifications. *Clin. Chem. Lab. Med.*, 2015, 53, p. 925-937.
7. **Westgard, J. O., Carey, R. N., Wold, S.** Criteria for judging precision and accuracy in method development and evaluation. *Clin. Chem.*, 1974, 20, p. 825–833.
8. https://www.westgard.com/biodatabase1.htm. Accessed October 2016.
9. **Oosterhuis, W. P, Theodorsson, E.** Total error vs. measurement uncertainty: revolution or evolution? *Clin. Chem. Lab. Med.*, 2016, 54, p. 235-239.
10. **Oosterhuis, W. P.** Gross overestimation of total allowable error based on biological variation. *Clin. Chem. Lab. Med.*, 2011, 57, p. 1334-1336.
11. **Henny, J., Vassault, A., Boursier, G., et al.** Recommendation for the review of biological reference intervals in medical laboratories. *Clin. Chem. Lab. Med.*, 2016 54, p. 1893-1900.
12. **Fraser, C. G.** Biological Variation: From Principles to Practice. *Washington, DC: ACC Press, 2001*
13. **Harry, M. J.** Mysteries of Six Sigma: Statistical Constructs and Engineering Rationale. *Phoenix: Palladyne Publishing; 2003.*
14. **Carobene, A.** Reliability of biological variation data available in an online database: need for improvement. *Clin. Chem. Lab. Med.*, 2015., 53, p. 871-877.
15. **Coskun, A., Oosterhuis, W. P., Serteser, M., Unsal, I.** Sigma metric or defects per million opportunities (DPMO): the performance of clinical laboratories should be evaluated by the Sigma metrics at decimal level with DPMOs. *Clin. Chem. Lab. Med.*, 2016, 54, p. 217-219.

16. **Budina, M.** Six Sigma- pověsti a fakta. *FONS*, 2016, 26, p. 17-21.
17. **Bais, R., Chester, D.** More on Lot-to-Lot Changes. *Clin. Chem.*, 2014, 60, p. 413-414. DOI: 10.1373/clinchem.2013.215111 Published January 2014
18. **Friedecký, B.** Variabilita mezi šaržemi reagencií a kalibrátorů a její důsledky pro verifikaci měření. *Klin. Biochem. Metab.*, 2016, 24, p. 170–172.

Do redakce došlo 27. 3. 2017

Adresa pro korespondenci:
RNDr. Peter Huba
Alpha Medical s.r.o,
Department of Clinical Chemistry and Haematology,
Školská 153/2
034 91 Lubochňa, Slovakia
e-mail: peter.huba@alphamedical.sk

Appendix

Appendix 1. Mathematical background of the S model

The reference interval is generally defined as an interval corresponding to 95% of the population centered on the median. In the case of normally distributed data, the reference interval corresponds to ±1.96 SD around the mean. The lower and upper limits of a reference interval in terms of the Z-score equal to -1.96 and +1.96, respectively. A Z-score is the number of standard deviations from the mean a data point is. The Z-score formula can be expressed as:

$$Z = \frac{(X - \mu)}{\sigma}$$

Where,
X = a normal random variable
μ = mean of a population
σ = standard deviation of a population

The width of a population reference interval (CV_{RI}), is defined mostly by biological variation and analytical variation of a method used to measure reference values. In an ideal situation, the analytical variation would be neglected and the width of a population RI becomes defined only by biological variation itself:

$$CV_{RI} = CV_B \leq \sqrt{(CV_I^2 + CV_G^2)}$$

Or

$$SD_{RI} = SD_B \leq \sqrt{(SD_I^2 + SD_G^2)}$$

The Z-score calculation will be as follows.

$$X = UL\ (LL) = 1.96 \times SD_B + \mu, \qquad (\mu - 1.96 \times SD_B)$$

$$Z = \frac{(X - \mu)}{\sigma} = \frac{1.96 \times SD_B + \mu - \mu}{SD_B} = 1.96$$

Analogously, for the lower limit the Z= -1.96

The Z-score of 1.96, with respect to the standardized normal distribution table, means the area under the normal curve equals to 0.975 (97.5%). In other words, the probability of a result to be greater than μ+1.96σ is 2.5%. Thus the ER_B of a reference interval defined only by the biological variation is 5%.

A maximum allowable imprecision

The S model considers the contribution of an analytical error that causes a maximum additional increase of 1.68% of patient results outside the original distribution. In such a situation, ER_B= 6.68% (ER_U= 3.34%) will be observed against expected 5%.

The calculation of maximum allowable CV_A will be as follows.

$$Z = \frac{(X - \mu)}{\sigma}$$

Where,

$$X = UL = 1.96 \times SD_B + \mu$$

$$SD_T \leq \sqrt{(SD_I^2 + SD_G^2 + SD_A^2)} = \sqrt{(SD_B^2 + SD_A^2)}$$

$$SD_B = \frac{CV_B \times \mu}{100} \ ; \ SD_T = \frac{CV_T \times \mu}{100}$$

$$Z = \frac{(X - \mu)}{\sigma} = \frac{\frac{1.96 \times CV_B \times \mu + (\mu - \mu)}{100}}{\frac{CV_T \times \mu}{100}} = \frac{1.96 \times CV_B}{CV_T}$$

$$Z = \frac{1.96 \times CV_B}{CV_T}$$

The Z-score of 1.833 means the area under the normal curve equals to 0.9666 (96.66%). In other words, the probability of a result to be greater than μ+1.96σ is 3.34%

$$CV_T = \frac{1.96 \times CV_B}{Z} = \frac{1.96 \times CV_B}{1.833} = 1.0693 CV_B$$

$$CV_T = \sqrt{(CV_B^2 + CV_A^2)}$$

$$1.0693 CV_B = \sqrt{(CV_B^2 + CV_A^2)}$$

$$1.0693^2 \times CV_B^2 = CV_B^2 + CV_A^2$$

$$1.1434 \times CV_B^2 - CV_B^2 = CV_A^2$$

$$0.1434 \times CV_B^2 = CV_A^2$$

$$\sqrt{(0.1434 \times CV_B^2)} = CV_A$$

$$CV_A = 0.379 \times CV_B$$

$$\frac{CV_A}{CV_B} = 0.379$$

A maximum allowable bias

The calculation of maximum allowable B_A will be as follows.

$$Z = \frac{(X - \mu)}{\sigma}$$

Where,

$$X = UL\ (LL) = 1.96 \times SD_B + \mu, \qquad (\mu - 1.96 \times SD_B)$$

$$\mu = \text{biased mean of a reference population} = \left(\frac{\mu \times B_A}{100}\right) + \mu, \ (B_A \text{ is in \%})$$

$$\sigma = SD_B = \text{standard deviation of reference values} = \frac{CV_B \times \mu}{100}$$

$$Z = \frac{(X - \mu)}{\sigma} = \frac{\frac{1.96 \times CV_B \times \mu}{100} + \mu - \left(\frac{\mu \times B_A}{100} + \mu\right)}{\frac{CV_B \times \mu}{100}} = 1.96 - \frac{B_A}{CV_B}$$

$$Z = 1.96 - \frac{B_A}{CV_B}$$

Since the CV_A, in this case, is assumed to be zero, the whole error budget = 0.379 x CV_B, falls on the bias.

Then,

$$\frac{B_A}{CV_B} = 0.379$$

The position of an upper limit of RI, against biased mean, will be

$$Z = 1.96 - 0.379 = 1.581$$

The Z-score of 1.581 means the area under the normal curve equals to 0.9431 (94.31%). In other words, the probability of a result to be greater than μ+1.96σ is 5.69%.

The position of a lower limit of RI, against biased mean, will be

$$Z = -1.96 - 0.379 = -2.339$$

The Z-score -2.339 means the area under the normal curve equals to 0.0097 (0.97%). In other words, the probability of a result to be lower than μ-1.96σ is 0.97%.

The sum of an error rate of the both sides of a distribution than equals to 6.66%.

A synergistic effect of bias and imprecision on bi-directional error rate, according to the S model

The calculation of a synergistic effect of the allowable B_A/CV_B and CV_A/CV_B will be as follows.

For illustration, consider this allocation of an analytical error:

$$\frac{B_A}{CV_B} = 0.286; \quad \frac{CV_A}{CV_B} = 0.249$$

$$EM_C = \sqrt{\left(\frac{B_A}{CV_B}\right)^2 + \left(\frac{CV_A}{CV_B}\right)^2}$$

$$EM_C = \sqrt{0.286^2 + 0.249^2} = 0.379$$

$$Z = \frac{(X - \mu)}{\sigma}$$

Where,

μ = biased mean of a reference population =

$$\left(\frac{\mu \times B_A}{100}\right) + \mu, \ (B_A \text{ is in } \%)$$

σ= SD_T=total standard deviation of test results in a reference population

$$SD_T \le \sqrt{(SD_I^2 + SD_G^2 + SD_A^2)} = \sqrt{(SD_B^2 + SD_A^2)}$$

$$SD_B = \frac{CV_B \times \mu}{100}; \ SD_T = \frac{CV_T \times \mu}{100}$$

The left tail of a distribution:

$$Z = \frac{(X - \mu)}{\sigma} = \frac{\mu - \frac{1.96 \times CV_B \times \mu}{100} - \left(\frac{\mu \times B_A}{100} + \mu\right)}{\frac{CV_T \times \mu}{100}} =$$

$$= \frac{-1.96 \times CV_B - B_A}{CV_T}$$

$$Z = \frac{-1.96 \times CV_B - B_A}{CV_T}$$

When,

$$B_A = 0.286 \times CV_B$$

$$CV_T = \sqrt{(CV_B^2 + CV_A^2)}$$

$$CV_A = 0.249 \times CV_B$$

Then,

$$Z = \frac{-1.96 \times CV_B - B_A}{CV_T} = \frac{-1.96 \times CV_B - 0.286 \times CV_B}{\sqrt{(CV_B^2 + (0.249 \times CV_B)^2)}} =$$

$$= \frac{-2.246 \times CV_B}{\sqrt{(1.062 \times CV_B^2)}}$$

$$Z = \frac{-2.246}{1.031} = -2.178$$

The Z-score of -2.178 means the area under the normal curve equals to 0.0147 (1.47%). In other words, the probability of a result to be lower than μ-1.96σ is 1.47%.

The right tail of a distribution:

$$Z = \frac{1.96 \times CV_B - B_A}{CV_T} = \frac{1.96 \times CV_B - 0.286 \times CV_B}{\sqrt{(CV_B^2 + (0.249 \times CV_B)^2)}} =$$

$$= \frac{1.674 \times CV_B}{\sqrt{(1.062 \times CV_B^2)}}$$

$$Z = \frac{1.674}{1.031} = 1.624$$

The Z-score of 1.624 means the area under the normal curve equals to 0.9478 (94.78%). In other words, the probability of a result to be greater than μ+1.96σ is 5.22%.

The sum of an error rate of the both sides of a distribution than equals to 6.69%.

Appendix 2. Mathematical background of the state-of-the-art approach

A maximum allowable imprecision

The width of a common reference interval could be expressed by the formula

$$CV_{RI} = \sqrt{(CV_I^2 + CV_G^2 + CV_{A0}^2)} = \sqrt{(CV_B^2 + CV_{A0}^2)}$$

Where CV_{A0} is the analytical variation of a method when the reference limits were determined. In the case that actual imprecision $CV_A \le CV_{A0}$ and bias can be neglected, then performance goal is defined by the CV_{A0}.

A maximum allowable bias

The maximum allowable bias will be as follows.

$$Z = \frac{(X - \mu)}{\sigma}$$

Where,

$X = UL\ (LL) = 1.96 \times SD_{RI} + \mu, \ (\mu - 1.96 \times SD_{RI})$

μ = biased mean of a reference population =

$$\left(\frac{\mu \times B_A}{100}\right) + \mu, \ (B_A \text{ is in } \%)$$

σ= SD_T=total standard deviation of test results in a reference population

$$SD_T \leq \sqrt{(SD_I^2 + SD_G^2 + SD_A^2)} = \sqrt{(SD_B^2 + SD_A^2)}$$

$$SD_{RI} = \frac{CV_{RI} \times \mu}{100};\ SD_T = \frac{CV_T \times \mu}{100}$$

$$Z = \frac{(X - \mu)}{\sigma} = \frac{\frac{1.96 \times CV_{RI} \times \mu}{100} + \mu - \left(\frac{\mu \times B_A}{100} + \mu\right)}{\frac{CV_T \times \mu}{100}} =$$

$$= \frac{1.96 \times CV_B - B_A}{CV_T}$$

Since the goal is to maintain the applicability of RI, the Z-score in the formula is replaced by 1.96. After modification, the maximum allowable B_A will be

$$B_A = 1.96 \times (CV_{RI} - CV_T)$$

Appendix 3. The CV_I/CV_g dependency of the conventional model

If we take a look at the example of, e.g., serum creatine and calcium listed below, we can find that the same quality specifications resulted from the conventional model, leads to the different error rate. These examples demonstrate an effect of CV_A, B_A on population RI, defined by the expression (5). The consequential ER_B is compared to expected ER_B= 5% (2.5%/ 2.5% below and above limit, respectively).

Serum creatinine (S-CREAT)

CV_I = 5.95, CV_G = 14.7%, CV_B = 15.86%, $\frac{CV_I}{CV_G}$ = 0.40

Serum calcium (S-CA)

CV_I = 2.1%, CV_G = 2.5%, CV_B = 3.26%, $\frac{CV_I}{CV_G}$ = 0.84

Condition 1 $CV_A = 0.5 \times CV_I$, $B_A = 0$

S-CREAT: ER_B= 5.42 %= + 0.42 % (2.71 % / 2.71 %)

S-CA: ER_B= 6.24 %= +1.24 % (3.12 % / 3.12 %)

Condition 2 $CV_A = 0$, $B_A = 0.25 \times CV_B$

Since the bias is derived from the CV_B, the effect on ER_B of population RI is the same regardless of CV_I/CV_g.

S-CREAT: ER_B= 5.72 %= + 0.72 % (1.36 % / 4.36 %)

S-CA: ER_B= 5.72 %= + 0.72 % (1.36 % / 4.36 %)

Condition 3 $CV_A = 0.5 \times CV_I$, $B_A = 0.25 \times CV_B$

S-CREAT: ER_B = 6.14 % = + 1.14 % (1.49 % / 4.65 %)

S-CA: ER_B = 7 % = + 2 % (1.79 % / 5.21 %)