MUSE

Multi Sample Evaluation

SKML Scoring
and
Reporting system

Version 2.8

November 2023

Contents

1. Introduction

The Multi Sample Evaluation (MUSE) system provides for systems for the case-oriented and the analyte-oriented quantitative and qualitative SKML schemes.

The standardization of the scoring systems emerges in a particular light now that a worldwide discussion is going on about harmonization [1]. An important role in this is played by External Quality Assessment Schemes, EQAS [2]. The SKML quantitative schemes belong to the best categories mentioned here, because where possible, they are based on commutable samples [3], which cover the clinically relevant concentration range, have assigned values based on reference methods [4], and have a scoring system with tolerance ranges based on the TEa concept [5,6].

The harmonization ambition of SKML was launched in the year 2000 under the name of Calibration 2000 and continues to be of importance in attaining named quality [7, 8, 9, 10, 11]. In the meantime the project has a sequel under the name of Calibration 2.000 and commences in the new version the co-development of reference methods and materials which are needed to accomplish the ambitions in harmonization and standardization.

Within the accreditation of laboratories according to ISO 15189 outcome parameters play an increasingly important role. Scores achieved in EQA schemes belong to the most important of these parameters. The MUSE reports aim to help you on the one hand to see the extent to which your corrective action is needed and on the other hand to bring into focus the extent to which your previous corrective actions were sufficiently effective.

The quantitative multi sample statistical approach in MUSE has been published in 2017. [12]

2. MUSE statistics for quantitative scoring systems

2.1. Commutable samples where possible

For a correct scoring system commutability of the samples used is essential. If commutability is demonstrated according to CLSI-C53-A or otherwise, this is mentioned in the description of the scheme concerned.

2.2. Target values

If possible reference values assigned by reference methods are used for determining the target values. If there are no reference values available, target values - determined by expert laboratories - or consensus method group averages are used. By which method value assignment has taken place, is mentioned in the description of the scheme concerned. Uncertainty in reference values and other target values is not yet included in the calculations for determining the width of the tolerance range.

2.3. Clinically relevant concentration range

As far as possible samples include the clinically relevant concentration range. The samples are selected on this basis. In different schemes target values are assigned to a low and a high (sometimes spiked) sample using reference methods in reference laboratories. By mixing these samples in different ratios, samples are obtained with intermediate and by calculation known concentrations.

Regression lines of laboratory results against target values (reference values, expert laboratory values or consensus method group averages) are time-weighted, with the most recent results receiving the greatest weight in the calculation of the regression line. Regression lines are calculated only if there are more than 3 results.

2.4. Tolerance ranges

Scores are assigned on the basis of two tolerance ranges, in which results must be located: the Total Error allowable (TEa) tolerance range, and the State of the Art (SA) tolerance range.

The SA tolerance range is a function of the concentration with a shape determined by the analytical precision profile. The SA tolerance range has a width m ± 3SDsa, whereby m is the target value (consensus method group average, or where available the value obtained by reference or expert laboratory). The SDsa is basically re-established every 3 years and calculated over a period of 6 years (see the procedure for calculating the precision profile in section 2.9).

The TEa tolerance range is also a function of the concentration and includes a range around the target values (reference values, expert laboratory values or consensus method group averages). The width of the tolerance range is a function of the concentration and is extrapolated from the concentration level at which the value of TEa is determined. This value is determined according to the EFLM Milan consensus criteria for analytical performance specifications. Conference criteria [13], whereby biological variation data take the most important place, unless clinical decision limits lead to another choice. The information used for biological variation shall be reviewed every year on the basis of the information available at www.westgard.com/biodatabase1.htm (until 1-1-2020) [6] and biologicalvariation.eu/meta_calculations [14] from 1-1-2020 onwards. The actual values for both SDsa and TEa are available on the SKML website at www.skml.nl/en/home/schemes/reportings/skml-tolerance-ranges.

2.5. Outlier removal

Average values and SDs are calculated per method group: methods which should lead to comparable results are classified in the same method group. To avoid disturbance of the result of the calculations by extreme and/or incorrect values (outliers), in the past outliers were removed on the basis of statistical tests. Later a technique was developed based on curve fitting. Since April 2022 robust statistics are used, in which NEN-ISO 13528:2015 (Statistical methods for use in proficiency testing by interlaboratory comparison) is leading. Robust statistics are even less sensitive to outliers and non-normal distribution of results.

NEN-ISO 13528:2015 provides many recommendations. A number of easy to calculate linear algorithms (L-estimators) have been chosen that perform reasonably well in all kinds of circumstances:

> The center of a set of results (mean) is approximated by taking the median.
>
> The dispersion within a set of results (SD) is approximated as follows:
>
> > if n = 2, then | result 1 – result 2 | / square root (2)
> > if n > 2, then MADe; if MADe = 0, then nIQR; if nIQR = 0, then the arithmetic SD

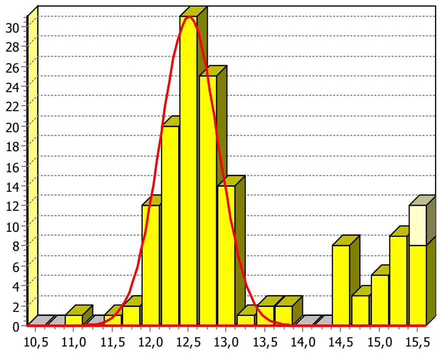
MADe stands for the scaled MAD (median absolute deviation) and nIQR stands for the normalized IQR (interquartile range). Outliers are any results that deviate more than 3 SD from the mean. If the SD has been determined arithmetically, then the outliers are removed and the mean and SD are re-approximated once without determining outliers again.

This calculation method is also carried out for instrument groups. Within the method and instrument groups the averages and SDs are calculated from the methods respectively instruments, omitting outliers. An average and SD is also calculated for all results, regardless of the method used. This is presented as ALTM (All Labs Trimmed Mean).

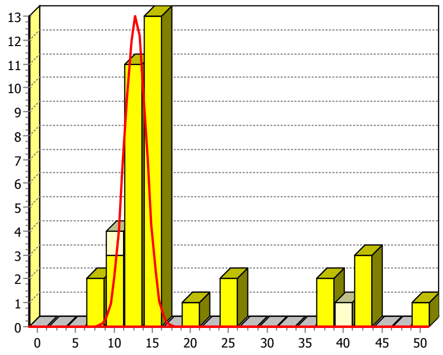
As well as the outlier removal on the basis of deviation with respect to the reference or consensus, outlier removal is also applied on the basis of deviation with respect to own laboratory. When an individual point has a significant other deviation (statistical as analytical) with respect to the own regression line, that result of that participant will be considered to be an outlier (see below paragraph 2.7: Within-laboratory SD)

Some examples:

PT



CA -19.9



2.6. Time weighting

The most recent values receive more weight in the calculations than results further back in time. The weighting parameters are calculated from:

$$W_i = 2^{-\Delta t/\propto}$$

Δt is the number of days (expressed in months) between the submission deadline of the last survey and the measurement date of sample i. The factor α is the half-life and is as standard 6 months, resulting in a weight for a result from a year before of 25% compared to the latest.

2.7. Within-laboratory SD

The within-laboratory SD is calculated as being the residual SD of the time-weighted regression line through the laboratory results versus the target values. A regression line (and therefore a within-laboratory SD) is calculated from a minimum of 4 and a maximum of 24 results. The regression line is forced through zero if the intersection does not deviate significantly from zero (Ttest at 95% reliability interval). If there are 4 measurement points, the line will always be forced through zero.

Outlier removal at individual laboratory level is done with the kSD method, where k is dependent on the number of measurements used in the regression calculation and the selected reliability, here 99,9%. If a result deviates statistically (more than k*SD) from the regression line, it is then tested whether the result also differs analytical relevant. For this purpose twice state of the art SDsa is used. If the result is considered as an outlier, the result will be excluded from the calculation of the within-laboratory SD, but will still receive a score.

If there are less than 4 points available for the regression line, the within lab precision is calculated according to:

Precision for sample concentration = target value for sample * VC at evaluation level (%) * root (ratio evaluation level/target value, where the ratio is never lower than 0.5)

The evaluation level is a concentration around the decision level of the relevant determination. This is often the concentration at which the biological variation is determined.

The average within-laboratory SD_{bl} of a consensus method group, method or instrument is calculated from all individual within-laboratory SD_{bli} according to:

$$SD_{bl} = \sqrt{\frac{\sum\left(SD_{bl_i}{}^{2} * (n_i - 1)\right)}{\sum (n_i - 1)}}$$

2.8. Between-laboratory SD

For every sample the between-laboratory SDs (SDtli) are calculated from the total SD (SDti) and the average within-laboratory SDbl at the concentration level of the sample. For this purpose, the average within-laboratory SD is extrapolated to the concentration level of the sample by extrapolation of the precision profile. Per sample the between-laboratory SDtli is calculated from the total SDti per sample and SDbl according to:

$$SD_{tli} = \sqrt{\left(SD_{ti}{}^{2} - SD_{bl}{}^{2}\right)}$$

From the between-laboratory SDtlis per sample is also calculated the between-laboratory SD per survey over all mi samples according to:

$$SD_{tl} = \sqrt{\frac{\sum\left(SD_{tli}{}^{2} * (m_i - 1)\right)}{\sum\left(m_i - 1\right)}}$$

2.9. Precision profiles

Precision profiles show the relationship between SD and concentration. Precision profiles are used to determine the shape of the TEa tolerance range and the SA tolerance range. In addition they are used to convert the within-laboratory SD to the different concentration levels of the samples used.

The precision profiles are calculated every 3 years, every time over the total SDs of all samples within a method in a period of 3 years. A practical calculation model has been chosen, as follows:

$$PP_t = \sqrt{\left[a^2 + \left(b\sqrt{X}\right)^2 + \left(cX\right)^2 + \left(\frac{d}{1 + X/e}\right)^2\right]}$$

The function contains four terms:

a) the axis-trim (constant SD)
b) a root term (constant variation)
c) a linear term (constant VC)
d) a hyperbole, describing the noise that often occurs in the low concentration range

The function proves to yield a good fit for almost all analytes.

A number of conditions apply to the curve fit, in order to obtain a valid function for each concentration. These conditions are:

1)	$d \geq \sqrt{(SD_{min}^2 - a^2)}$	d is such that if X = 0 the function is at least equal to the lowest measured SD
2)	$e \geq X_{min} / 2$	at the point at $X = X_{min} / 2$ the value of the hyperbole halves; X_{min} is the lowest measured concentration

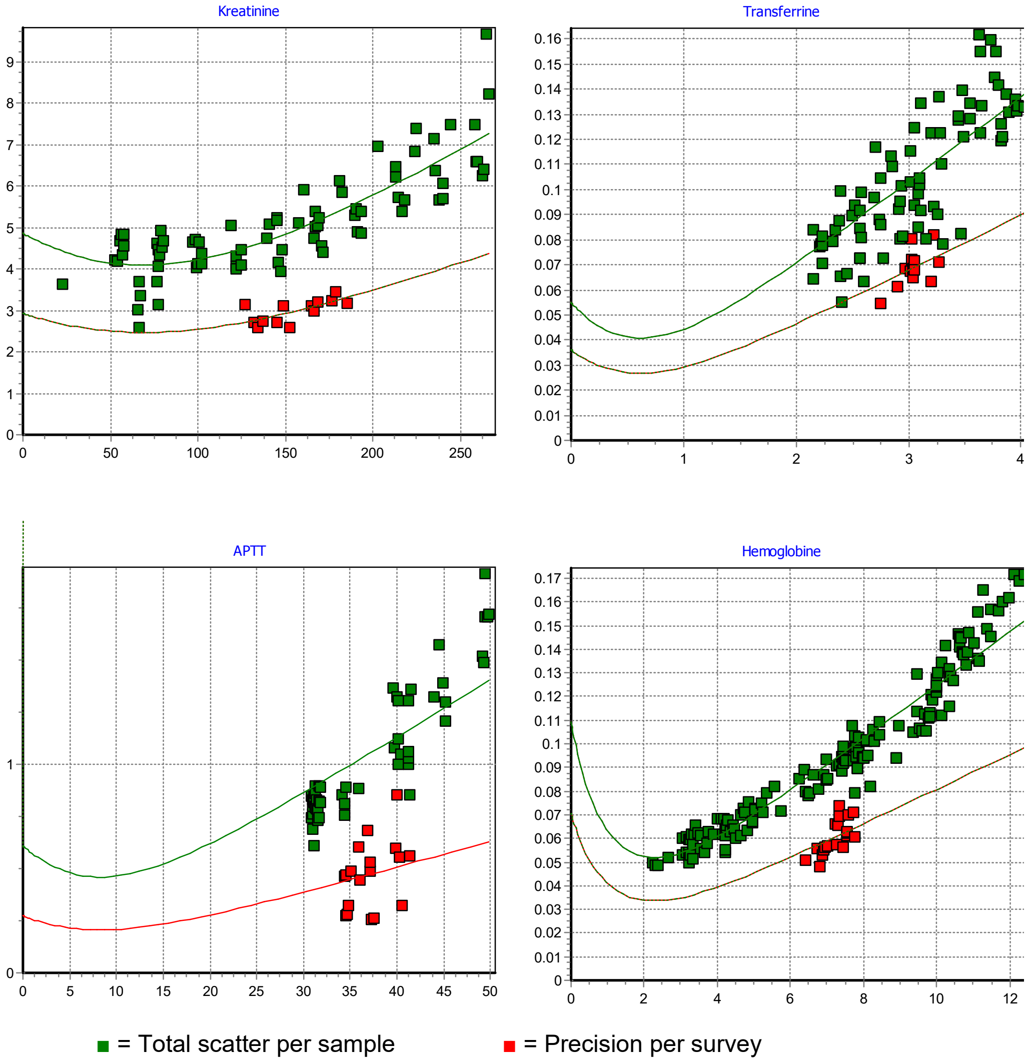
The precision profiles for the within-lab SD are set up on the basis of estimates of within-laboratory SDs from the same period versus the consensus method group averages of the concerning surveys. Through these points a curve is fitted with the shape:

$$PP_b = f * PP_t$$

Typical values for f are between 0.2 and 0.8.

To do justice to the designation 'state of the art' the section can potentially exclude methods which are of insufficient quality from these calculations when calculating the precision profiles.

Examples of precision profiles:



A precision profile can not be set in new schemes. In that case, the old "square root formula" can be used that determines the relationship between SDsa and concentration using the following formula:

$$VCsa = VCtarget * \sqrt{\frac{\text{evaluation level}}{x}}$$

When x < 2* evaluation level And

$$VCsa = VC\text{evaluation level} * \sqrt{\frac{1}{2}}$$

When x > 2* evaluation level

3. MUSE Scoring system for quantitative analytes

3.1. Performance score and Six Sigma

The scientific fundament of the MUSE scoring system is the theory of Six-Sigma. This is used world-wide to quantify the quality of a production process. In a process that meets the requirements of the Six-Sigma standard, the scatter is so low that less than 1:1000000 products do not meet the quality standard. Therefore, it is necessary that the scatter in the process is maximally (less than) 1/6 of the standard. Conversely the quality of a process can be expressed as the number of SDs (= sigma) that has been achieved. Therefore, if the SD is 1/3 of the standard, we speak of a 3-Sigma process, which translates into 0.2% rejection (= 99.8% complies).

Because the SKML thinks the requirement of 1:1000000 is too severe, a “satisfactory” standard of 95% adequate has been chosen, corresponding to a sigma of 2. A sigma of 4.5 comprises an “excellent” score.

The differences between the laboratory results and the targets eventually constitute the basis for the performance scores. For both TE and SA sigma values are calculated for each measurement according to the following formula:

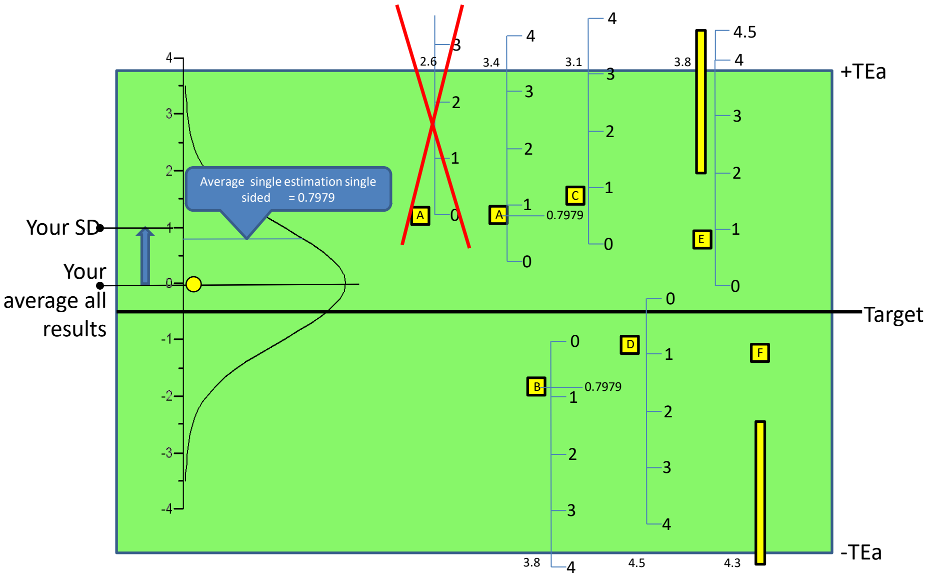
$$sigma_i = \frac{TL - |X_i - T_i|}{SD_{bl}} + \sqrt{2/\pi}$$

Here TL is the tolerance limit (TEa or SA); Xi the measured value and Ti the target for the sample (reference where possible, otherwise method group consensus). The term √2/π (≈ 0.7979) is a correction for the fact that sigma is not calculated from the average bias of the participant, but from a single estimate of that bias. If the individual measuring points would be measured in a high multiple, then the scale would start at the average of those points. Because not a multiple, but a single estimate is measured, we therefore must correct for the average location of the single measurement with respect to that average. This correction causes a shift of √2/π in the direction of the eccentricity on the spot.

The precision of the yellow bar is calculated as follows:

The time-weighted regression line from the submitted results versus the target values is used to calculate the indoor laboratory SD (imprecision). The SD is then calculated from the deviations from the individual measurements of the regression line.

In the figure below this calculation is shown graphically:

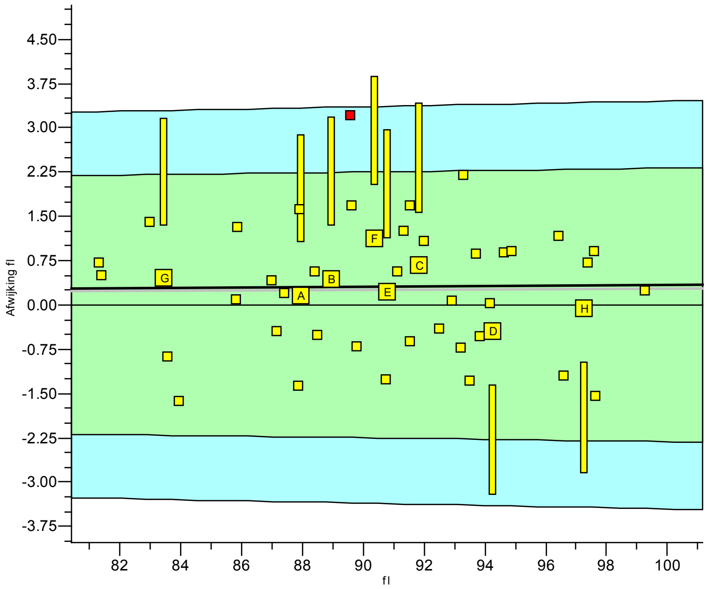


Displayed on the left is the expected distribution of the measurement results of the participant.

In this diagram, for samples A to F the calculated sigma values are corrected repeatedly with this factor and read on the edge of the green tolerance range. For samples A to E the sigma scales are shown; for sample E and F the sigma interval of 2 – 4.5 is marked using a yellow bar. With sample F the sigma scale is omitted, as is the case in the difference plot. Thus at a glance, per measurement result, it can be established whether the individual measurement meets the 2-sigma respectively the 4.5 sigma criterion.

The sigma values are calculated for all measurements and averaged for both the TEa tolerance range (sigma-TE) and the SA tolerance range (sigma-SA). Then, an average sigma with corresponding score is calculated: a value for sigma ≥2 is worth 1 point; a value of ≥4.5 is worth 2 points.

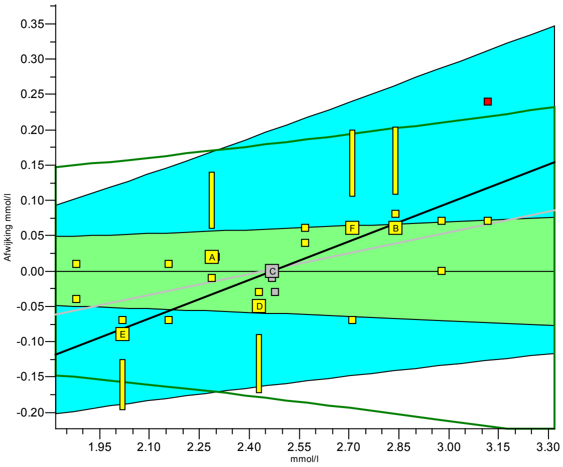
The example below is a difference plot of the MCV of one of the participants. The TEa respectively the SA tolerance ranges are displayed with a green respectively blue colour. The measurement results of the participant are the yellow cubes. The 8 measurement results of the current survey are marked with a letter. It is easily visible that all yellow bars fall partially in the green TEa range, hence a TE-score of one point. Because only 2 measurement results fall outside the blue SA tolerance range to a small extent, there is a SA score of two points. The difference plot is discussed in more detail in paragraph 5.1.3.



Finally a Performance score (P-score) is assigned.

A score is only assigned if the target is either based on a reference value, or on an expert value or consensus based on ALTM. This means that no scores are assigned if the target value is method-group dependent. In those cases participants can judge their performance on the basis of the sigma values if they indeed intend to measure in accordance to the method group. If a laboratory deliberately has calibrated its method in accordance to a different method group, the laboratory cannot use sigma values to judge its performance, and has to rely on graphical comparison of its results to the intended method in the histograms for each sample.

The score is either on the basis of the sigma TE, based on the biological variation concept, or on the basis of the sigma SA based on the State of the Art (SA) tolerance range. The section determines per analyte whether sigma TE or sigma SA is used. When the blue SA range is wider than the green TEa range, then the precision to be achieved according to the TEa concept is apparently larger than possible according to the current state of the art. In those cases the section will decide in principle to score on SA. With available reference or expert method the desired trueness remains determined by the position of the green TEa range. Hence in cases where the SA tolerance range is wider than the TEa tolerance range the latter is widened to the width of the SA tolerance range. This is displayed graphically in the difference plot by means of green lines, which mark the outside of the corrected TEa tolerance range. Below an example:



The maximum Performance score (P-score) per investigation is 2. The Performance scores are set as standard:

Sigma	P-score
≥ 4.5	2
2.0 - 4.5	1
< 2.0	0

The section may decide to assign negative Performance scores to some analytes when a TE sigma value is considered to be so small that this could lead to wrongful treatment with serious consequences. The minimum Performance score is then -1 or -2. As standard then applies:

Sigma	P-score
≥ 4.5	2
2.0 - 4.5	1
1.0 - 2.0	0
0-1.0	-1
<0	-2

Although negative Performance scores are assimilated into the cumulative score, they are always recorded separately in red.

The Performance scores of all analytes reported in a survey within a cluster, are totalised to the Survey Performance score (Survey P score).

The current trueness is the average deviation of the samples of this survey from the target values. The cumulative trueness is the time-weighted average deviation over the time frame of the scheme.

The current precision is the within-laboratory SD. This is also the cumulative precision because the within-laboratory SD is calculated from the time-weighted regression line over the time frame of the scheme.

3.2. Maximum Achievable Performance (MAP)

The maximum P-score per analyte is 2 corresponding with a sigma of >4.5. The MAP is then equal to the number of analytes from this measurement environment times 2.

4. MUSE Scoring system for qualitative studies

4.1. Clinically relevant concentration range or composition

Samples must cover the clinically relevant concentration range as much as possible or have the clinically relevant composition, like the combination of micro-organisms and density of a parasite.

4.2. Performance score

The coordinator decides if and to which studies within a survey Performance scores are assigned. The coordinator assigns the achievable Performance score points on the basis of expert findings or consensus results. The maximum Performance score per investigation or conclusion question is 2. The minimum Performance score is -2. Points are assigned by the coordinator of the scheme as follows:

- The maximum Performance score of 2 is assigned to investigation results or answers to conclusion questions which completely correspond to the reference or expert result or, where not available, the consensus result.
- A score of 1 is assigned to investigation results or answers to conclusion questions which partially, or at least to a sufficient extent, correspond to the expert result or, where not available, the consensus result.
- Whenever the reported result is “elsewhere” (sent to), a score of 0 from 0 will be given. The percentage of answers reported “elsewhere” is mentioned separately in the review.
- A score of 0 is assigned to investigation results or answers to conclusion questions which do not correspond to the expert result or, where not available, the consensus result.
- A negative Performance score is assigned to investigation results or answers to conclusion questions which may result in wrongful diagnosis or treatment:
  - A score of -2 has been reserved for wrong results or answers to conclusion questions which may result in treatments or lack thereof with very serious or fatal complications.
  - A score of -1 has been reserved for wrong results or answers to conclusion questions which may result in wrongful treatment, but with limited complications.

Performance scores are totalized by sample as well as survey level. Negative scores are emphasized by means of a red colour.
Like in quantitative determinations scores are only assigned if the target value is not dependent on the method used.

4.3. Case-oriented versus Non case-oriented

In the assessment of (the total of) the scores it is important to differentiate between case-oriented surveys and non-case-oriented surveys. In the case-oriented survey the participants are considered to have completed all results and answered all questions. Failure to provide a result or answer yields a score of 0 points, whilst the MAP remains the same. To accommodate participants who do not implement an analyte, or cannot answer a question, in selection lists an additional option is offered “Elsewhere” or an option of equal scope. Such a response is given a score of 0 from 0, it is replaced in the report by an empty result in the form of an empty square (□). However, the percentage of answers that is forwarded, is mentioned separately in the review.

The section can choose to add additional options which give a further indication why no analysis has been carried out, for example “Not relevant”.

Alternatively, the coordinator may decide to score in a non-case-oriented way. In that case the MAP is reduced proportionally if not all results/answers are given. The number of points for say the result “Elsewhere” will then usually be 2. Probably most schemes will be scored on the basis of a non-case-oriented way. The fact that a survey has one or more case studies, certainly does not mean that the survey will be scored case-oriented.

4.4. Maximum Achievable Performance (MAP)

The MAP (Maximum Achievable Number of Points) is the sum of the maximum achievable number of points per result/question. In calculating the MAP a distinction is made between case-oriented and non-case-oriented:

- Case-oriented. MAP is the arithmetical sum of all maximum achievable number of points per result/question.
- Non-case-oriented. MAP is the sum of the maximum achievable number of points, for which the participant has reported a result.

5. General reporting design

MUSE uses a modular reporting system. The report consists of a number of modules which display dependent on the structure of the scheme. These modules are built around a number of graphic elements:

- Histogram
- Difference plot
- Score pictogram
- Score indicators

The following modules may be distinguished:

- Survey summary page with total scores
- Case-oriented reporting per sample
  - Sample information
  - Casuistry
  - Questioning
  - Determination
  - Qualitative and quantitative investigation results per analyte (list)
  - Conclusion questions
  - Histogram determination results
- Summary page with quantitative investigation results per analyte
- Summary page with scores quantitative investigation results per analyte (list)
- Summary page with scores qualitative investigation results per analyte (list)
- Summary page with scores for conclusion questions
- Analyte-oriented reporting (graphs)
  - Difference plot
  - Histograms quantitative results
  - Histograms qualitative results

5.1. Graphic elements

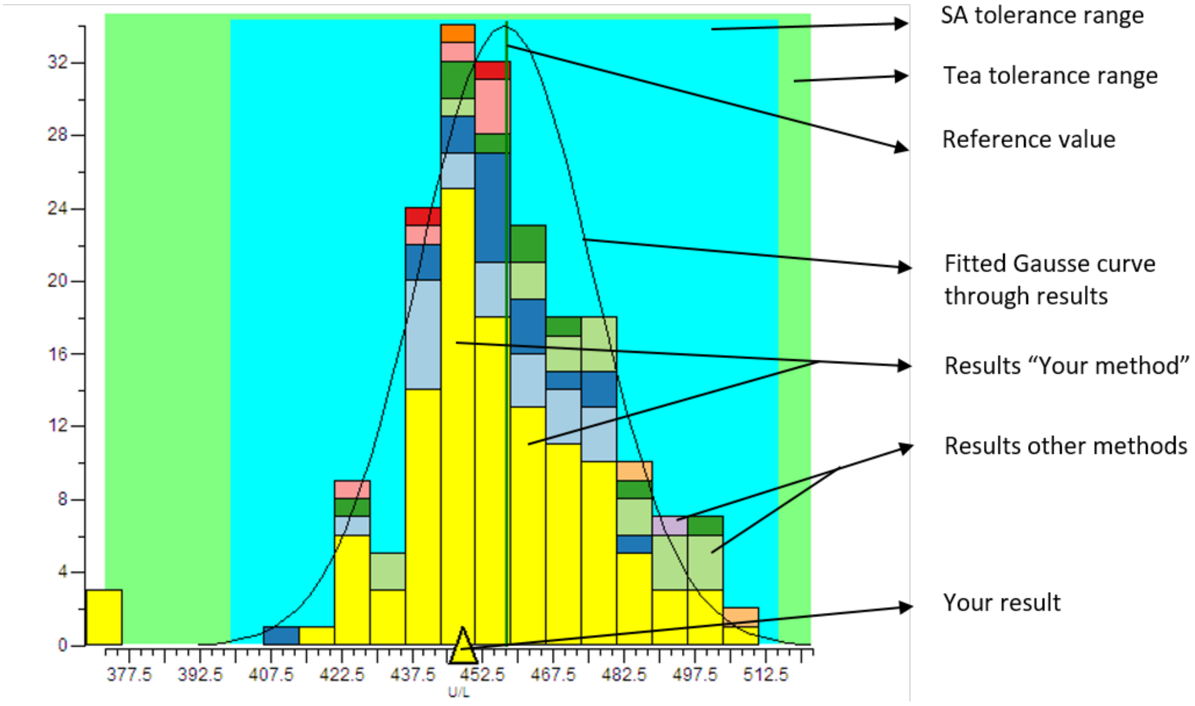
5.1.1. Use of colour

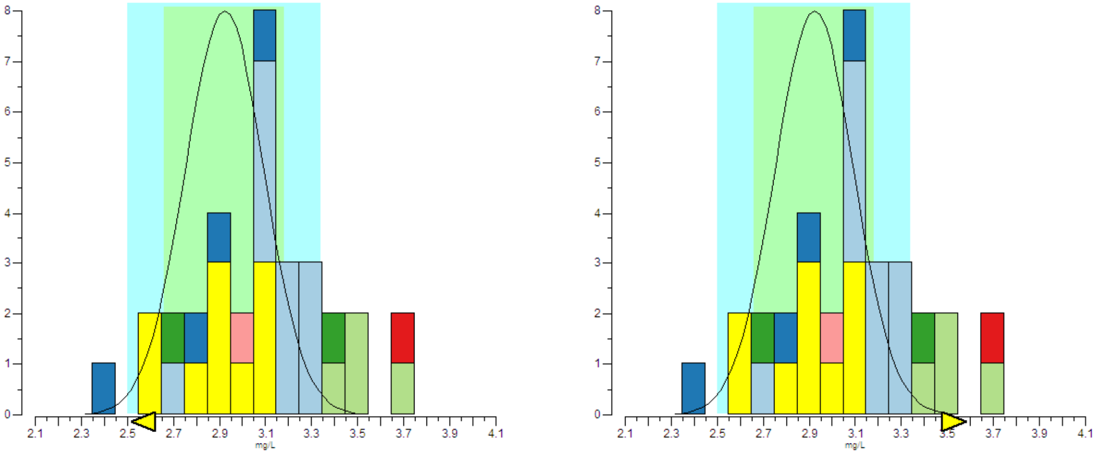
By using colours systematically, an attempt is made to quickly provide insight in the meaning of the various graphic components. The following colour scheme is used:

Colour	Meaning
🟨	Own result ; own method ; own score …..
🟧	Indication of method within own method group
🟦	SA Tolerance range; For this analyte reference values are used
🟩	TEa Tolerance range; For this analyte reference values are used
🟦	SA Tolerance range; For this analyte **no** reference values are used
🟩	TEa Tolerance range; For this analyte **no** reference values are used
🟥	Outlier
⬛	Result excluded from calculations
🟩	Reference value ; Expert value ; Weighed value

5.1.2. Histogram

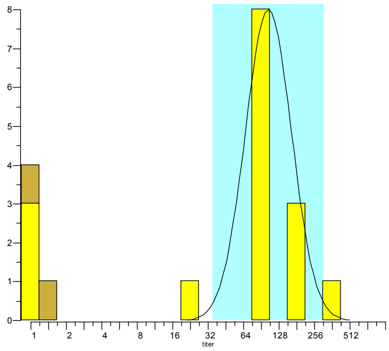
The histogram shows the distribution of all results across the concentration range. The height of each bar indicates the number of participants which have measured the result which is on the x-axis. The colour indicates the method used. The legend at the bottom of the histograms indicates how methods are assigned to colours. Yellow is always used for the own method. A upward pointing triangular yellow indicator at the x-axis indicates the own result. In cases of ‘less than’ or ‘greater than’ results, the indicator points respectively to the left or right rather than upwards.

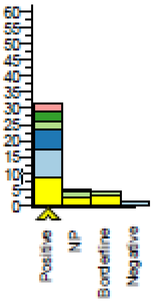


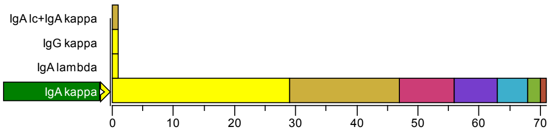


The left picture shows a participant’s histogram with a “less than” result, the right chart of a participant with a “greater than” result.

Variants:







Logarithmic x-as

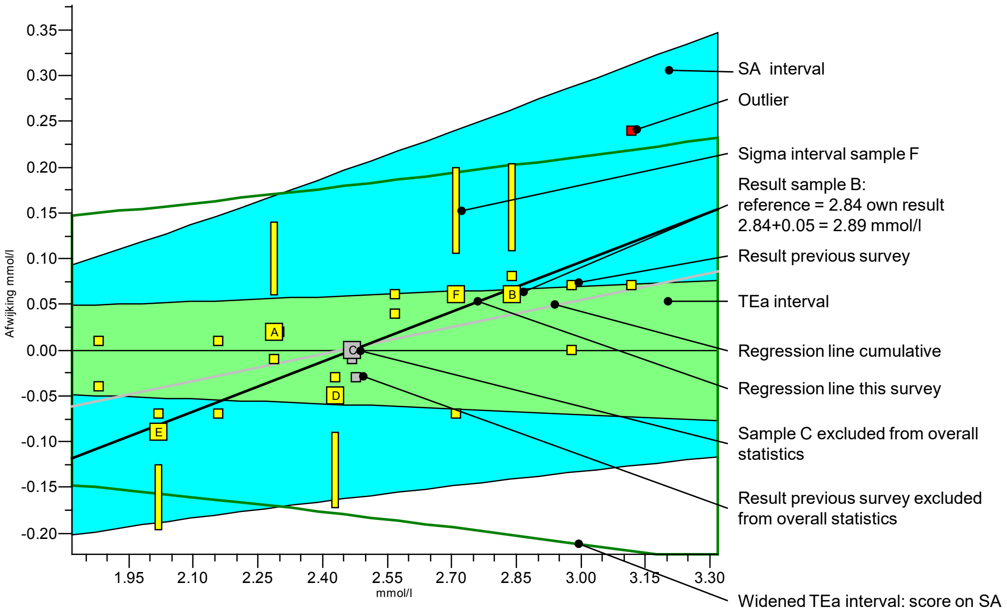
Qualitative histogram

Deterministic histogram

5.1.3. Difference plot

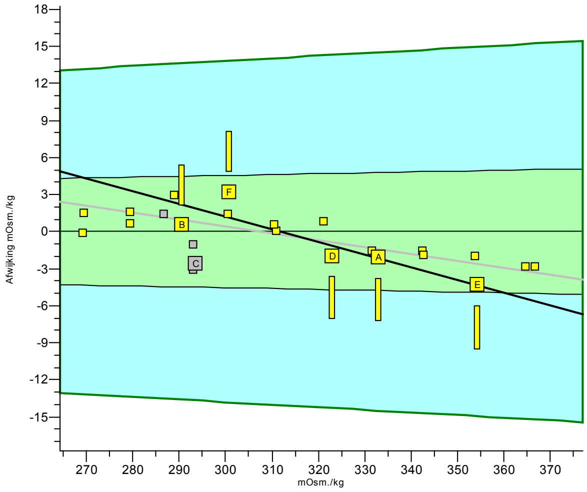
In the difference plot the results of a participant are shown as a function of the target value (reference if available, otherwise consensus value). On the y-axis is always the difference between the own result and the target value.

In the background the tolerance intervals for TEa and SA are shown and where necessary the widened to SA TEa tolerance. The results from the current survey are shown by means of yellow squares containing the letter of the concerning sample. Results from previous surveys are represented by smaller yellow squares.

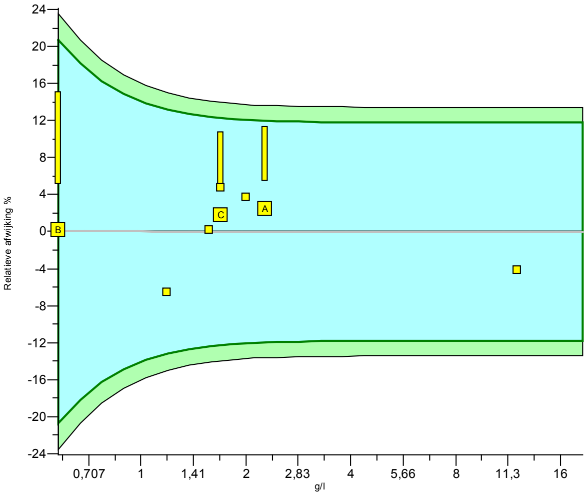


In the difference plot 2 regression lines are shown. These are time-weighted and calculated from the results achieved in this survey (black line) and cumulative (grey line). The regression lines are calculated after removal of outliers, in addition the regression line is forced through zero if the intersection does not deviate significantly from zero (Ttest at 95% confidence interval). If there are 4 measurement points, the line will always be forced through zero. the within-lab scatter is calculated as the residual scatter of the results around these regression lines.

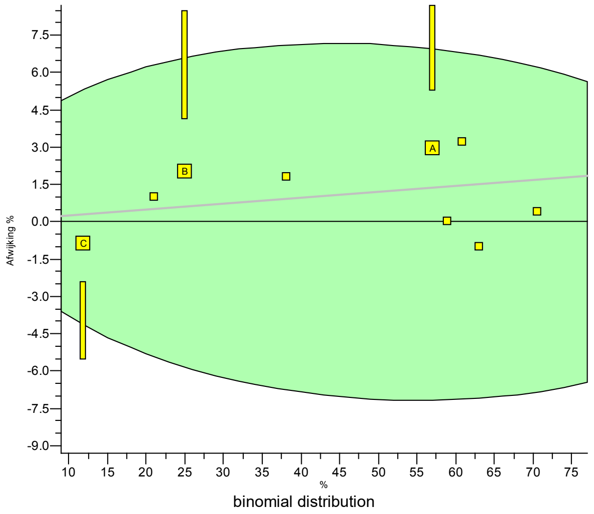
Variants:



Soft colours: no hard reference or expert values

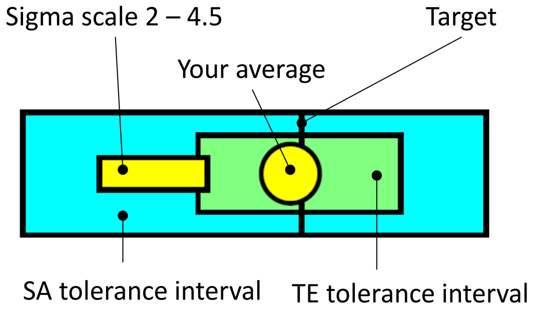


Logarithmic x-axis and a relative y-axis



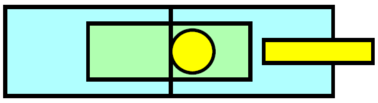
5.1.4. Score pictogram

The score pictogram is a thumbnail summary of the difference plot. It attempts to provide insight at a glance into the performance of the analyte concerned.

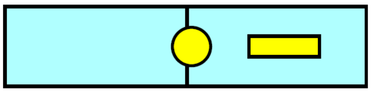


The green and blue squares symbolise again the TEa tolerance range and the SA tolerance range. The location of the average value of this participant is represented by the yellow circle and is therefore a representation of the eccentricity. The yellow bar is the sigma scale (2-4.5), which can be read from the edges of the TEa and SA tolerance ranges. Here it is obvious that the sigma-TE is slightly more than 2 (1 point) and the sigma-SA > 4.5 (2 points).

Variants:



Soft colours: no hard reference or expert values



No green area: there are no TEa standards set for this analyte

5.1.5. Score indicators

By using score indicators adequate and inadequate scores are shown. These indicators are used anywhere where scores are assigned, so for both quantitative analytes as qualitative and for determinations and conclusion questions. For each assigned score a cube is shown, whose colour determines the result:

🟩 = Adequate score (1 or 2 points)
🟥 = Inadequate score (<0 points)
□ = No result submitted, does not count for MAP

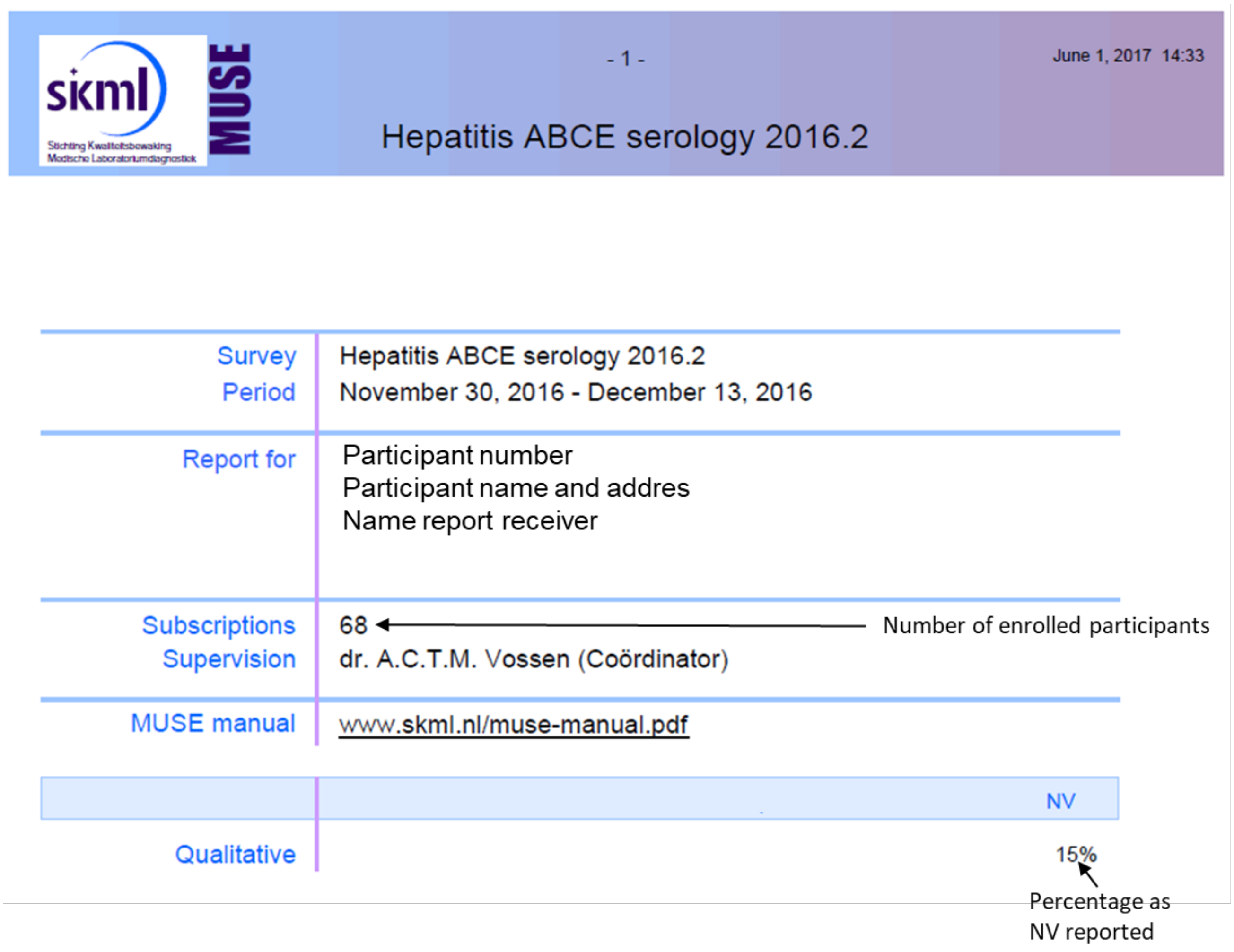
Example:

🟩🟩🟥🟩🟩🟩

1st, 2nd, 4th, 5th and 6th result correctly measured, 3rd result is incorrect.

5.2. Survey summary page with total scores

This page displays both general information of this survey and a review of the scores achieved by this survey. It depends on the arrangement of the survey which sub scores are shown here.



- 1 -

June 1, 2017 14:33

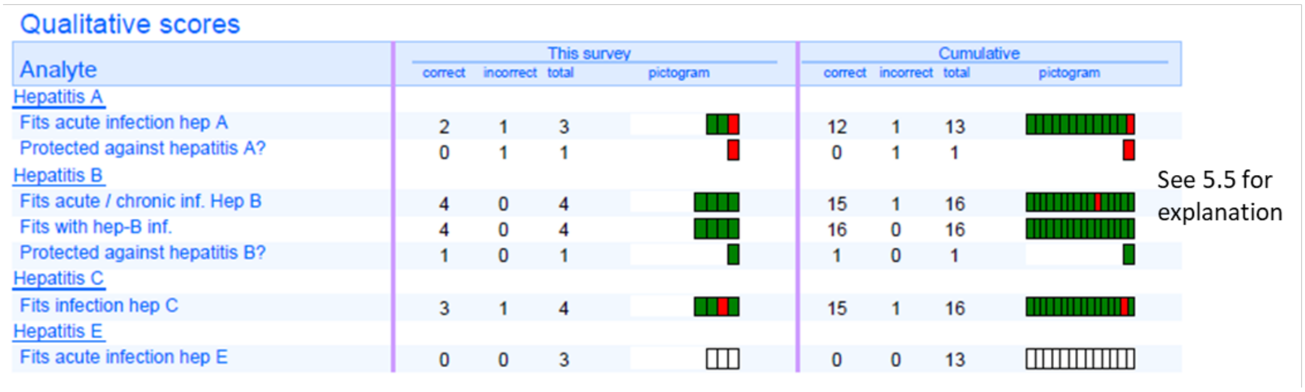
Hepatitis ABCE serology 2016.2

Survey Period	Hepatitis ABCE serology 2016.2 November 30, 2016 - December 13, 2016
Report for	Participant number Participant name and addres Name report receiver
Subscriptions Supervision	68 dr. A.C.T.M. Vossen (Coördinator)
MUSE manual	www.skml.nl/muse-manual.pdf

	NV
Qualitative	15%

Number of enrolled participants

Percentage as NV reported



Qualitative scores

Analyte	This survey				Cumulative			
	correct	incorrect	total	pictogram	correct	incorrect	total	pictogram
Hepatitis A								
Fits acute infection hep A	2	1	3	🟩🟩🟥	12	1	13	🟩🟩🟩🟩🟩🟩🟩🟩🟩🟩🟩🟩🟥
Protected against hepatitis A?	0	1	1	🟥	0	1	1	🟥
Hepatitis B								
Fits acute / chronic inf. Hep B	4	0	4	🟩🟩🟩🟩	15	1	16	🟩🟩🟩🟩🟩🟩🟩🟩🟩🟩🟥🟩🟩🟩🟩🟩
Fits with hep-B inf.	4	0	4	🟩🟩🟩🟩	16	0	16	🟩🟩🟩🟩🟩🟩🟩🟩🟩🟩🟩🟩🟩🟩🟩🟩
Protected against hepatitis B?	1	0	1	🟩	1	0	1	🟩
Hepatitis C								
Fits infection hep C	3	1	4	🟩🟩🟥🟩	15	1	16	🟩🟩🟩🟩🟩🟩🟩🟩🟩🟩🟩🟩🟩🟩🟥🟩
Hepatitis E								
Fits acute infection hep E	0	0	3	⬜⬜⬜	0	0	13	⬜⬜⬜⬜⬜⬜⬜⬜⬜⬜⬜⬜⬜

See 5.5 for explanation

5.3. Case-oriented reporting per sample

The MUSE reporting per sample is for all reporting layouts always the same and consists of four optional modules:

- Case study
- Results (qualitative and quantitative)
- Determinations
- Conclusion questions

5.3.1. *Modules Case study and results per sample*

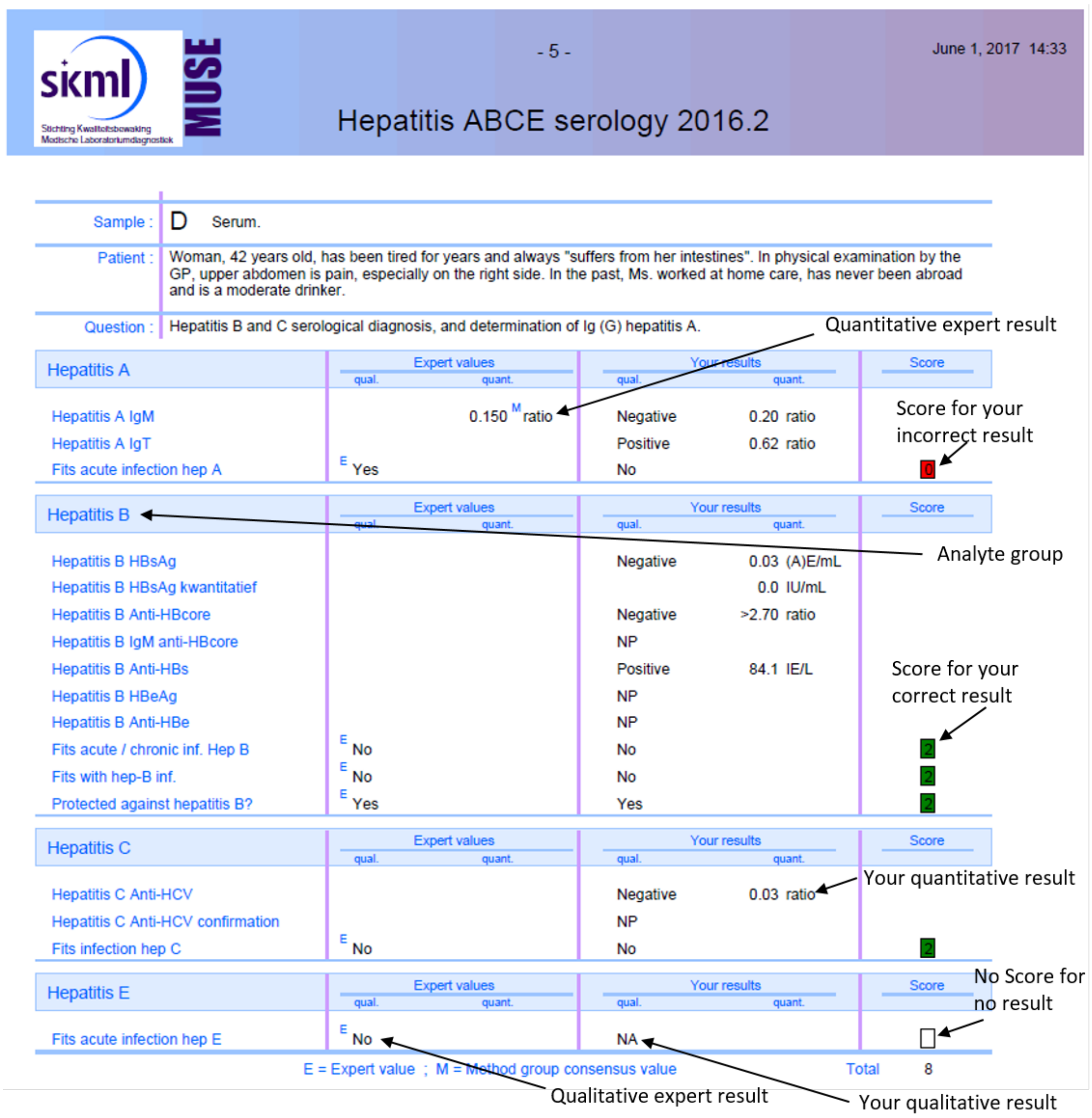
This module always starts with the case study as shown in QBase on the results screen. Per analyte group (not in every survey are the analytes divided into groups) a list follows with results and the associated expert values if assigned.

In the column “Expert values” qualitative and quantitative expert values can be incorporated, of which the origin is indicated by a letter and is explained in the legend:

R = Reference value. A value determined by a reference technique
E = Expert value. A value determined by the expert.
C = Consensus value. A qualitative value determined by the expert as in majority reported result
M = Method group consensus value. The average value of all results determined by “Your method group”

In the column “Score” the score for the qualitative result is displayed by a score indicator (see also paragraph 5.1.5):

2	= adequate score (1 or 2 points)
0	= inadequate score (>0 points)
□	= no result submitted, result does not count for MAP
<empty>	= no scores assigned to this analyte



- 5 -

June 1, 2017 14:33

Hepatitis ABCE serology 2016.2

Sample : D Serum.

Patient : Woman, 42 years old, has been tired for years and always "suffers from her intestines". In physical examination by the GP, upper abdomen is pain, especially on the right side. In the past, Ms. worked at home care, has never been abroad and is a moderate drinker.

Question : Hepatitis B and C serological diagnosis, and determination of Ig (G) hepatitis A.

Hepatitis A	Expert values		Your results		Score
	qual.	quant.	qual.	quant.	
Hepatitis A IgM		0.150 ^{M} ratio	Negative	0.20 ratio	
Hepatitis A IgT			Positive	0.62 ratio	
Fits acute infection hep A	^{E} Yes		No		0

Hepatitis B	Expert values		Your results		Score
	qual.	quant.	qual.	quant.	
Hepatitis B HBsAg			Negative	0.03 (A)E/mL	
Hepatitis B HBsAg kwantitatief				0.0 IU/mL	
Hepatitis B Anti-HBcore			Negative	>2.70 ratio	
Hepatitis B IgM anti-HBcore			NP		
Hepatitis B Anti-HBs			Positive	84.1 IE/L	
Hepatitis B HBeAg			NP		
Hepatitis B Anti-HBe			NP		
Fits acute / chronic inf. Hep B	^{E} No		No		2
Fits with hep-B inf.	^{E} No		No		2
Protected against hepatitis B?	^{E} Yes		Yes		2

Hepatitis C	Expert values		Your results		Score
	qual.	quant.	qual.	quant.	
Hepatitis C Anti-HCV			Negative	0.03 ratio	
Hepatitis C Anti-HCV confirmation			NP		
Fits infection hep C	^{E} No		No		2

Hepatitis E	Expert values		Your results		Score
	qual.	quant.	qual.	quant.	
Fits acute infection hep E	^{E} No		NA		□

E = Expert value ; M = Method group consensus value Total 8

Quantitative expert result

Score for your incorrect result

Analyte group

Score for your correct result

Your quantitative result

No Score for no result

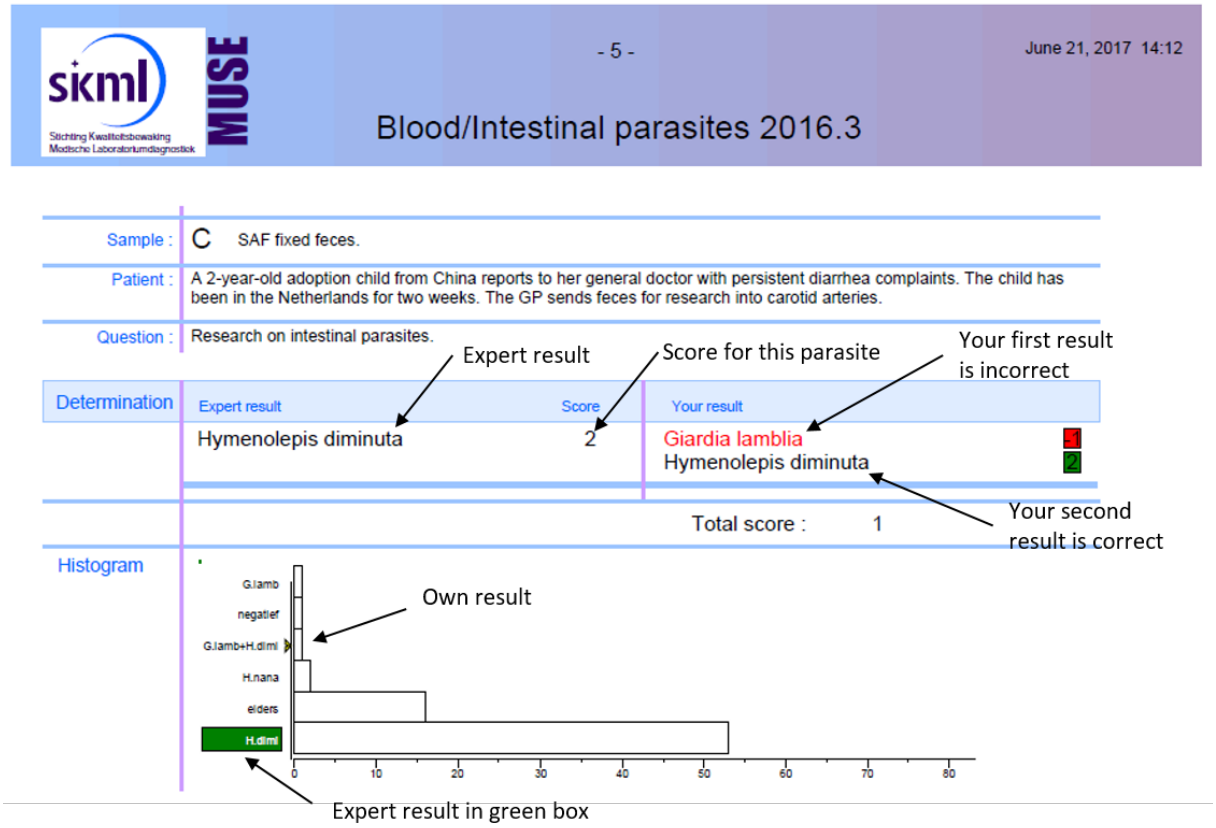
Qualitative expert result

Your qualitative result

5.3.2. Module Determination

This module always follows the case study. In the left column the obtainable micro-organisms are shown and the number of achievable points. In the right column are the results of the participant and behind that, indicated by a green or red indicator (see paragraph 5.1.5) whether the parasite found corresponds to the expert result.

Below the list a histogram is displayed containing the combinations of parasites found. The expert result is indicated by a green bar and the own result by a yellow arrow.



Stichting Kwaliteitsbewaking Medische Laboratoriumdiagnostiek

- 5 -

June 21, 2017 14:12

Blood/Intestinal parasites 2016.3

Sample : C SAF fixed feces.

Patient : A 2-year-old adoption child from China reports to her general doctor with persistent diarrhea complaints. The child has been in the Netherlands for two weeks. The GP sends feces for research into carotid arteries.

Question : Research on intestinal parasites.

Determination

Expert result	Score	Your result	
Hymenolepis diminuta	2	Giardia lamblia	🟥 -1
		Hymenolepis diminuta	🟩 2

Total score : 1

Histogram

Expert result

Score for this parasite

Your first result is incorrect

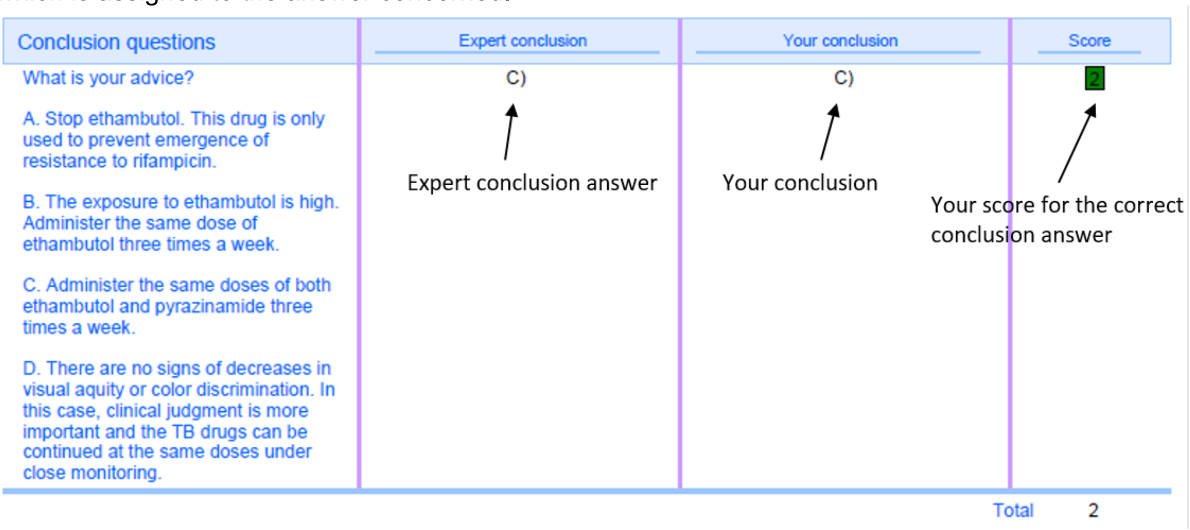
Your second result is correct

Own result

Expert result in green box

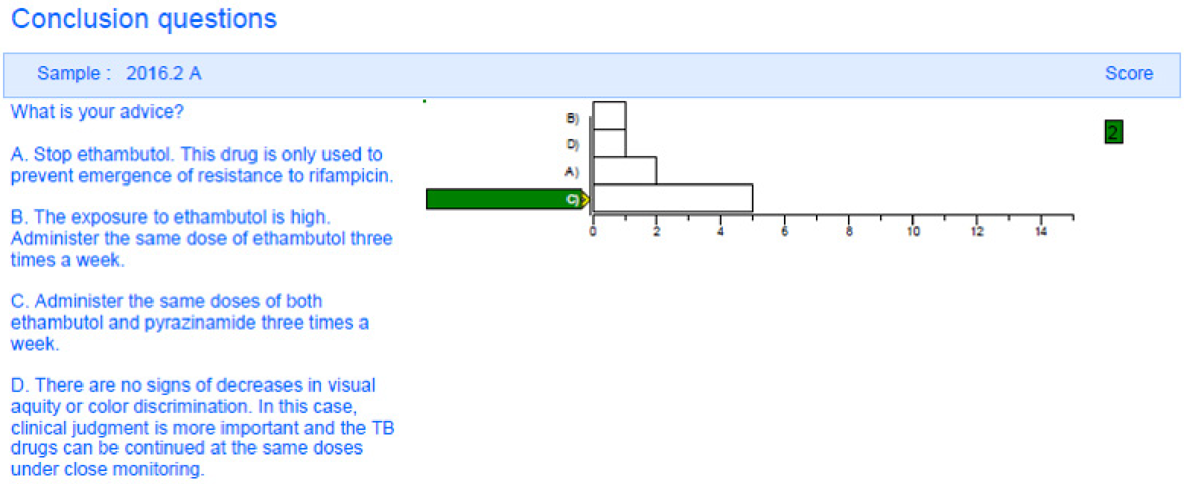
5.3.3. Module Conclusion questions

The conclusion questions are shown after the qualitative results for the sample concerned (and the possible determinations). Left is the expert conclusion, right the own conclusion. Far right is the score which is assigned to the answer concerned.



Conclusion questions	Expert conclusion	Your conclusion	Score
What is your advice? A. Stop ethambutol. This drug is only used to prevent emergence of resistance to rifampicin. B. The exposure to ethambutol is high. Administer the same dose of ethambutol three times a week. C. Administer the same doses of both ethambutol and pyrazinamide three times a week. D. There are no signs of decreases in visual aquity or color discrimination. In this case, clinical judgment is more important and the TB drugs can be continued at the same doses under close monitoring.	C)	C)	2
		Total	2

At the end of the report a summary is given of the distribution of the answers to the conclusion questions:

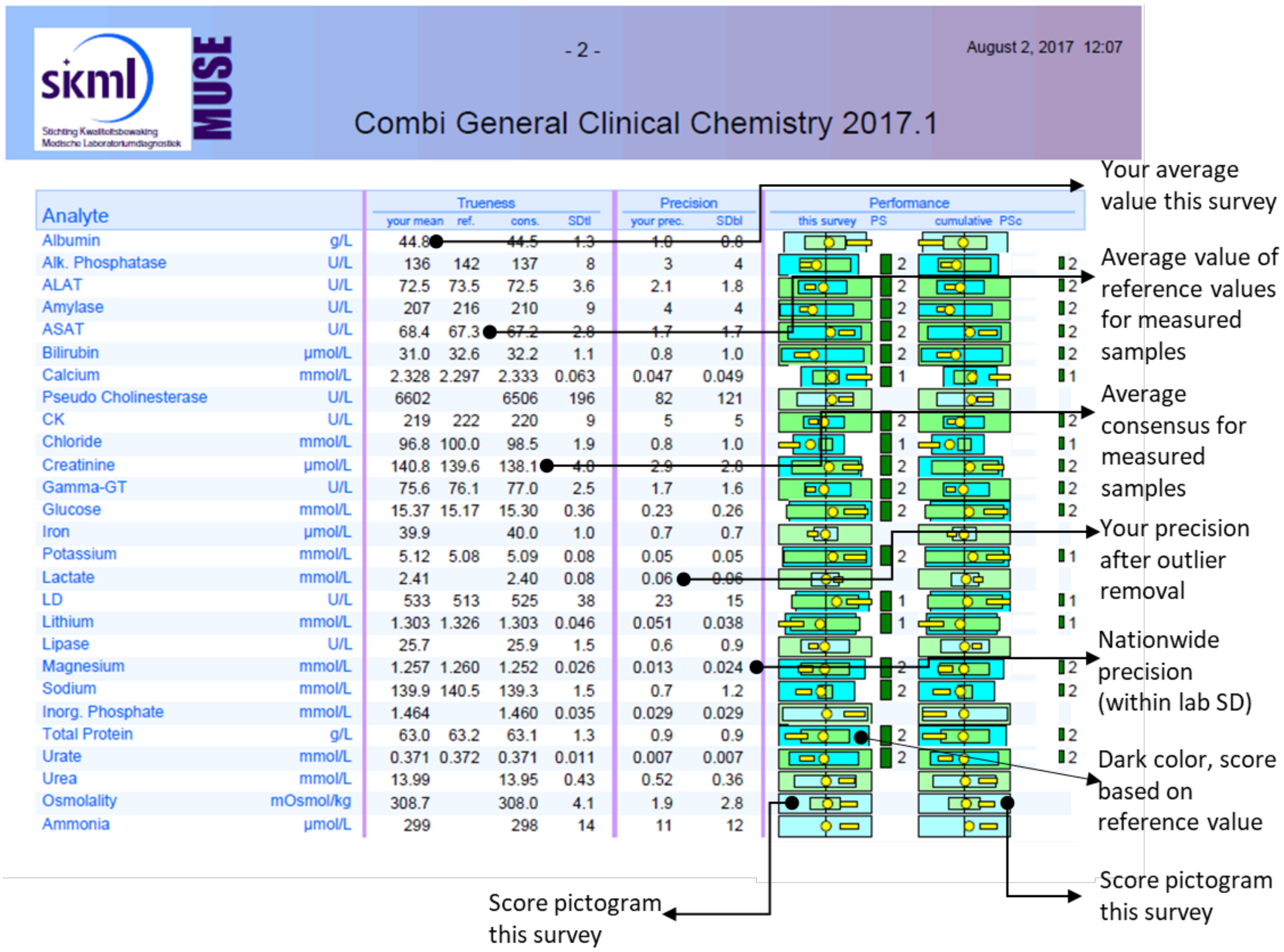


Conclusion questions

Sample : 2016.2 A	Score
What is your advice? A. Stop ethambutol. This drug is only used to prevent emergence of resistance to rifampicin. B. The exposure to ethambutol is high. Administer the same dose of ethambutol three times a week. C. Administer the same doses of both ethambutol and pyrazinamide three times a week. D. There are no signs of decreases in visual aquity or color discrimination. In this case, clinical judgment is more important and the TB drugs can be continued at the same doses under close monitoring.	2

5.4. Quantitative summary page

On this page the different quantitative parameters for trueness and precision are shown per analyte. Also on this page the performance scores for both this survey and cumulative are represented by the score pictogram and the score itself.



- 2 -

August 2, 2017 12:07

Combi General Clinical Chemistry 2017.1

Analyte		Trueness				Precision		Performance			
		your mean	ref.	cons.	SDtl	your prec.	SDbl	this survey	PS	cumulative	PSc
Albumin	g/L	44.8		44.5	1.3	1.0	0.8				
Alk. Phosphatase	U/L	136	142	137	8	3	4		2		2
ALAT	U/L	72.5	73.5	72.5	3.6	2.1	1.8		2		2
Amylase	U/L	207	216	210	9	4	4		2		2
ASAT	U/L	68.4	67.3	67.2	2.8	1.7	1.7		2		2
Bilirubin	µmol/L	31.0	32.6	32.2	1.1	0.8	1.0		2		2
Calcium	mmol/L	2.328	2.297	2.333	0.063	0.047	0.049		1		1
Pseudo Cholinesterase	U/L	6602		6506	196	82	121				
CK	U/L	219	222	220	9	5	5		2		2
Chloride	mmol/L	96.8	100.0	98.5	1.9	0.8	1.0		1		1
Creatinine	µmol/L	140.8	139.6	138.1	4.0	2.9	2.8		2		2
Gamma-GT	U/L	75.6	76.1	77.0	2.5	1.7	1.6		2		2
Glucose	mmol/L	15.37	15.17	15.30	0.36	0.23	0.26		2		2
Iron	µmol/L	39.9		40.0	1.0	0.7	0.7				
Potassium	mmol/L	5.12	5.08	5.09	0.08	0.05	0.05		2		1
Lactate	mmol/L	2.41		2.40	0.08	0.06	0.06				
LD	U/L	533	513	525	38	23	15		1		1
Lithium	mmol/L	1.303	1.326	1.303	0.046	0.051	0.038		1		1
Lipase	U/L	25.7		25.9	1.5	0.6	0.9				
Magnesium	mmol/L	1.257	1.260	1.252	0.026	0.013	0.024		2		2
Sodium	mmol/L	139.9	140.5	139.3	1.5	0.7	1.2		2		2
Inorg. Phosphate	mmol/L	1.464		1.460	0.035	0.029	0.029				
Total Protein	g/L	63.0	63.2	63.1	1.3	0.9	0.9		2		2
Urate	mmol/L	0.371	0.372	0.371	0.011	0.007	0.007		2		2
Urea	mmol/L	13.99		13.95	0.43	0.52	0.36				
Osmolality	mOsmol/kg	308.7		308.0	4.1	1.9	2.8				
Ammonia	µmol/L	299		298	14	11	12				

Your average value this survey

Average value of reference values for measured samples

Average consensus for measured samples

Your precision after outlier removal

Nationwide precision (within lab SD)

Dark color, score based on reference value

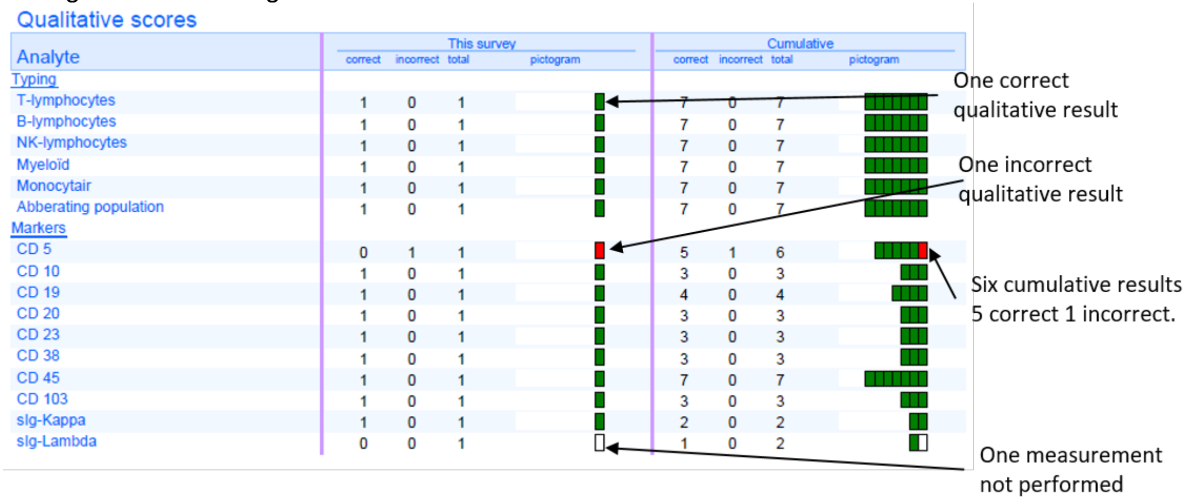
Score pictogram this survey

Score pictogram this survey

The values are calculated from time-weighted individual underlying results. In the calculation of the reference - expert - and consensus values only those points are involved for which the participant has submitted a result which is not classified as outlier. Hence the respective average may vary between participants and clusters!

5.5. Qualitative score page

On this page score indicators (see paragraph 5.1.5) are used to give a review of the scores achieved for the qualitative results, for both the current survey and cumulative. Through red and green colours respectively the number of incorrect and (almost) correct results are counted. Results which do score but are not reported by the participant, are indicated by an empty box. The results are chronologically arranged from left to right.



Qualitative scores

Analyte	This survey				Cumulative			
	correct	incorrect	total	pictogram	correct	incorrect	total	pictogram
Typing								
T-lymphocytes	1	0	1	🟩	7	0	7	🟩🟩🟩🟩🟩🟩🟩
B-lymphocytes	1	0	1	🟩	7	0	7	🟩🟩🟩🟩🟩🟩🟩
NK-lymphocytes	1	0	1	🟩	7	0	7	🟩🟩🟩🟩🟩🟩🟩
Myeloïd	1	0	1	🟩	7	0	7	🟩🟩🟩🟩🟩🟩🟩
Monocytair	1	0	1	🟩	7	0	7	🟩🟩🟩🟩🟩🟩🟩
Abberating population	1	0	1	🟩	7	0	7	🟩🟩🟩🟩🟩🟩🟩
Markers								
CD 5	0	1	1	🟥	5	1	6	🟩🟩🟩🟩🟩🟥
CD 10	1	0	1	🟩	3	0	3	🟩🟩🟩
CD 19	1	0	1	🟩	4	0	4	🟩🟩🟩🟩
CD 20	1	0	1	🟩	3	0	3	🟩🟩🟩
CD 23	1	0	1	🟩	3	0	3	🟩🟩🟩
CD 38	1	0	1	🟩	3	0	3	🟩🟩🟩
CD 45	1	0	1	🟩	7	0	7	🟩🟩🟩🟩🟩🟩🟩
CD 103	1	0	1	🟩	3	0	3	🟩🟩🟩
sIg-Kappa	1	0	1	🟩	2	0	2	🟩🟩
sIg-Lambda	0	0	1	□	1	0	2	🟩□

One correct qualitative result

One incorrect qualitative result

Six cumulative results 5 correct 1 incorrect.

One measurement not performed

5.6. Analyte-oriented reporting

In this report section the quantitative and qualitative results per analyte are shown in 3 modules:

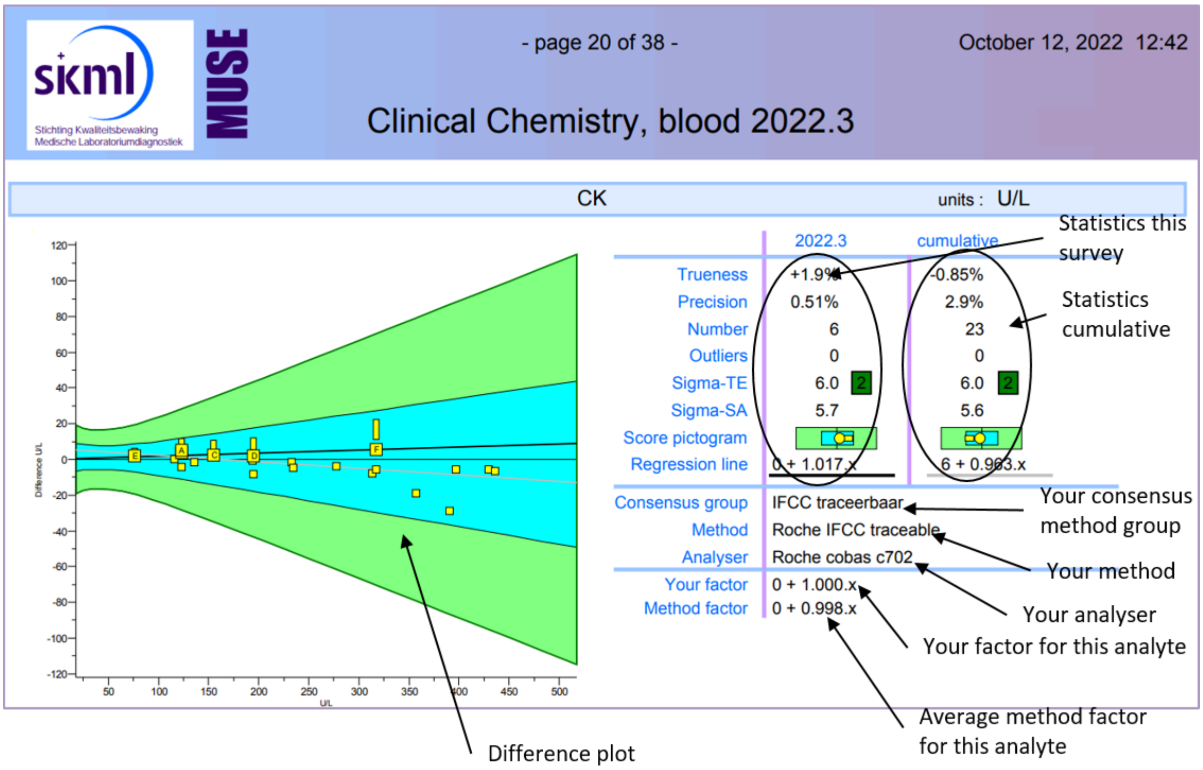
- Difference plot
- Histograms quantitative
- Histograms qualitative

5.6.1. Module difference plot

In this module the results of the current survey and the previous surveys are shown in the form of a difference plot. How many results are shown, depends on the horizon determined for this scheme by the section, usually this is 1 year.

Left is the difference plot itself and right the related statistical data, for both this survey and the cumulative statistics, calculated over the surveys which fall within the horizon of this scheme. For the calculation of the regression line(s), trueness and precision the individual measuring points are processed time-weighted (see paragraph 2.7).

We recommend evaluating your performance based on the cumulative statistics and scores. MUSE is designed to make a robust statement about your bias and imprecision over a wide concentration range, based on enough observations. The statistics and scores for the survey are based on fewer points and therefore more sensitive to random effects on a sample. These statistics give a more turbulent picture, with a greater chance of unfavorable scores. However, we also provide these short-term statistics to enable you to quickly recognize changes that have occurred recently. This can be useful for changes in your method, for example based on a corrective action following a previous survey.



Stichting Kwaliteitsbewaking Medische Laboratoriumdiagnostiek

MUSE

- page 20 of 38 -

October 12, 2022 12:42

Clinical Chemistry, blood 2022.3

CK units : U/L

	2022.3	cumulative
Trueness	+1.9%	-0.85%
Precision	0.51%	2.9%
Number	6	23
Outliers	0	0
Sigma-TE	6.0 2	6.0 2
Sigma-SA	5.7	5.6
Score pictogram		
Regression line	0 + 1.017.x	6 + 0.963.x

Consensus group	IFCC traceerbaar
Method	Roche IFCC traceable
Analyser	Roche cobas c702
Your factor	0 + 1.000.x
Method factor	0 + 0.998.x

5.6.2. Module histograms quantitative

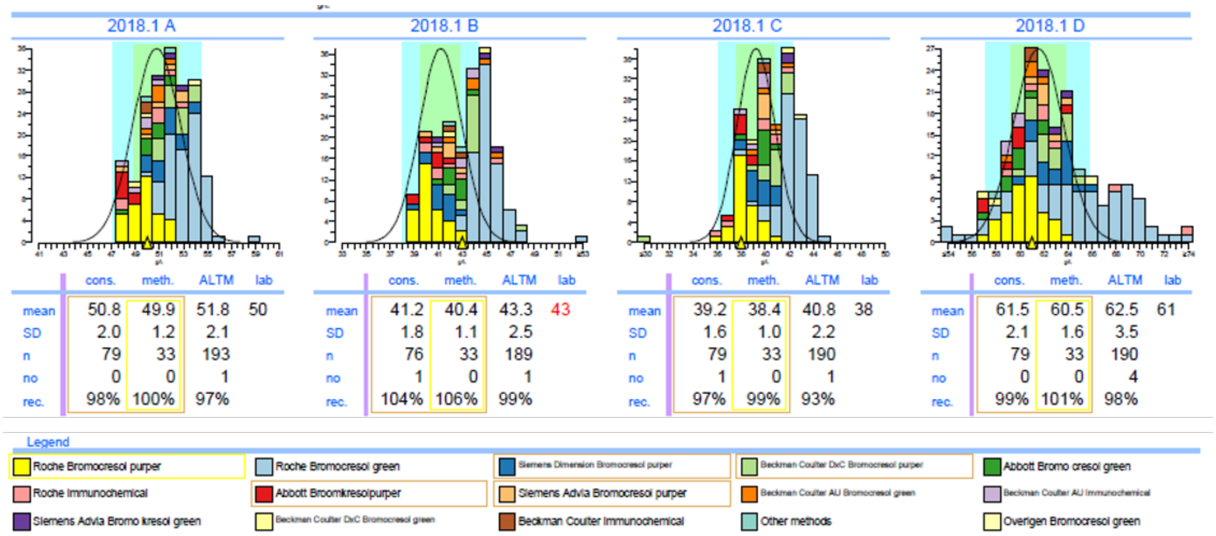
In this module the results of this survey are shown in the form of a histogram per sample. If the number of samples per survey is 1 or 2, the results of previous surveys are also shown to a maximum of 4.

The colours in the bars match the technique used and is explained in the legend.

In addition the own method is always displayed in yellow.

In the column ALTM (= All Labs Trimmed Mean) the average value is shown across all techniques, after removal of outliers. This gives an impression of the total population of submitted results. If there is a reference value or expert value, this is shown there.

Under the histograms of quantitative schemes, in addition to the average (mean), the standard deviation (SD), the number of results (n) and the number of outliers (now), you will also find the recovery (rec.). At rec. we show your result as a percentage of the upper row "gem. (average) which contains (method group) consensus, method average, and reference, expert, or ALTM, respectively.



2018.1 A

	cons.	meth.	ALTM	lab
mean	50.8	49.9	51.8	50
SD	2.0	1.2	2.1	
n	79	33	193	
no	0	0	1	
rec.	98%	100%	97%	

2018.1 B

	cons.	meth.	ALTM	lab
mean	41.2	40.4	43.3	43
SD	1.8	1.1	2.5	
n	76	33	189	
no	1	0	1	
rec.	104%	106%	99%	

2018.1 C

	cons.	meth.	ALTM	lab
mean	39.2	38.4	40.8	38
SD	1.6	1.0	2.2	
n	79	33	190	
no	1	0	1	
rec.	97%	99%	93%	

2018.1 D

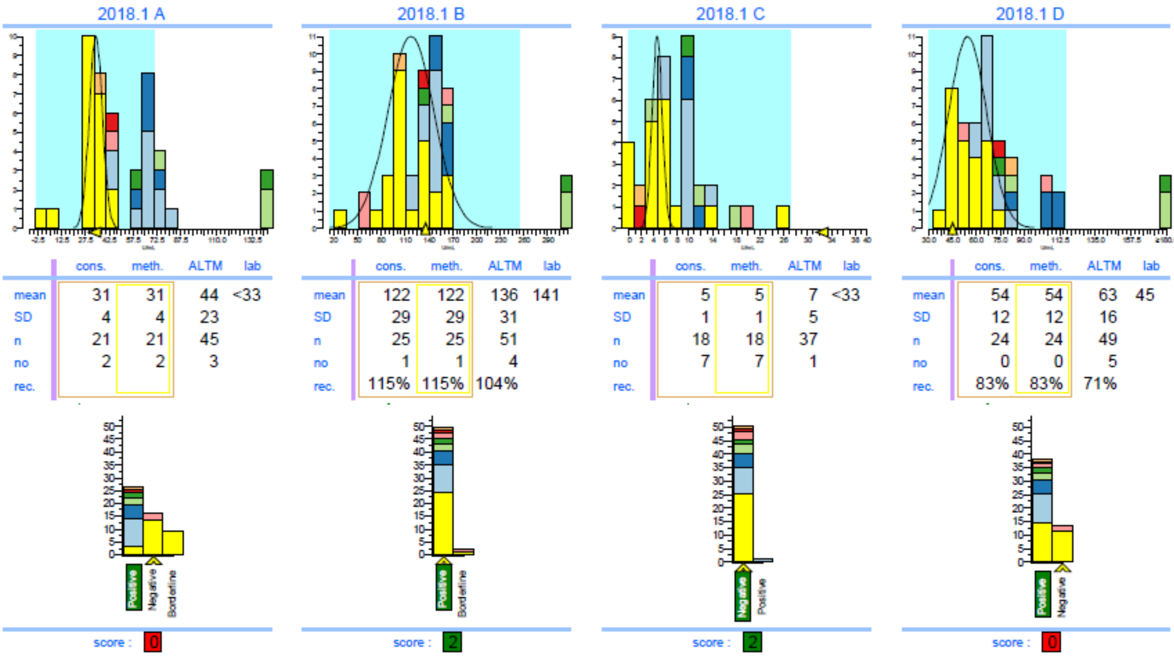
	cons.	meth.	ALTM	lab
mean	61.5	60.5	62.5	61
SD	2.1	1.6	3.5	
n	79	33	190	
no	0	0	4	
rec.	99%	101%	98%	

Legend

Roche Bromocresol purper	Roche Bromocresol green	Siemens Dimension Bromocresol purper	Beckman Coulter DxC Bromocresol purper	Abbott Bromo cresol green
Roche Immunochemical	Abbott Broomkresolpurper	Siemens Advia Bromocresol purper	Beckman Coulter AU Bromocresol green	Beckman Coulter AU Immunochemical
Siemens Advia Bromo kresol green	Beckman Coulter DxC Bromocresol green	Beckman Coulter Immunochemical	Other methods	Overigen Bromocresol green

5.6.3. Module histograms qualitative

The qualitative histograms (when available) are shown directly below the quantitative histograms (likewise when available). The use of colour is the same as in the quantitative histograms, so that the results can be linked to each other.



2018.1 A

	cons.	meth.	ALTM	lab
mean	31	31	44	<33
SD	4	4	23	
n	21	21	45	
no	2	2	3	
rec.				

score : 0

2018.1 B

	cons.	meth.	ALTM	lab
mean	122	122	136	141
SD	29	29	31	
n	25	25	51	
no	1	1	4	
rec.	115%	115%	104%	

score : 2

2018.1 C

	cons.	meth.	ALTM	lab
mean	5	5	7	<33
SD	1	1	5	
n	18	18	37	
no	7	7	1	
rec.				

score : 2

2018.1 D

	cons.	meth.	ALTM	lab
mean	54	54	63	45
SD	12	12	16	
n	24	24	49	
no	0	0	5	
rec.	83%	83%	71%	

score : 0

6. Annual report

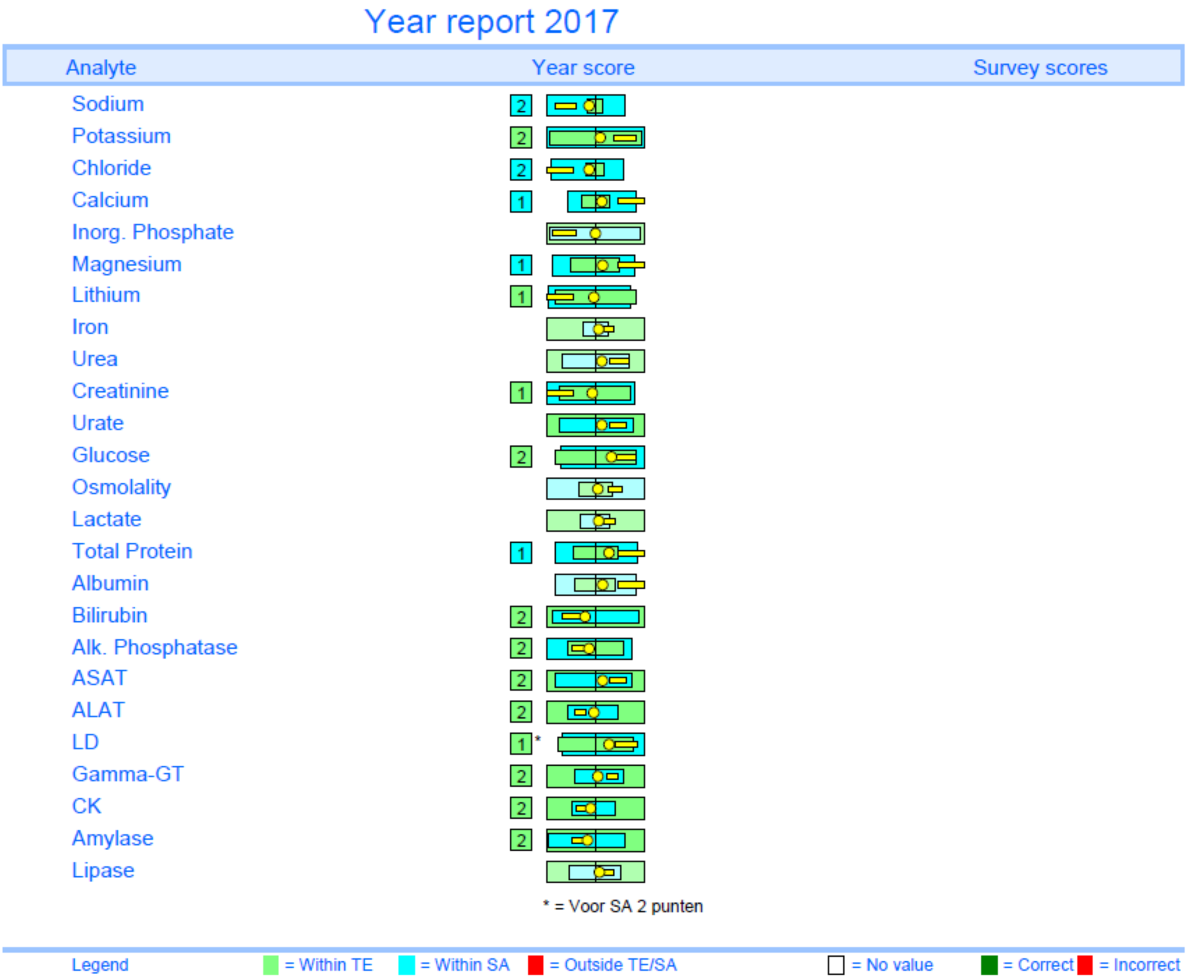
In the Annual report the cumulative results for each laboratory are shown per analyte, aggregated over all submitted clusters (if appropriate). As with the regular report, a distinction is made between quantitative, qualitative, deterministic and conclusion questions. An important difference in the quantitative annual report with the regular report is that in the annual report no use is made of time-dependent weighting factor.

6.1. Quantitative annual report

In the column “Annual score” the achieved annual score is shown per analyte, framed by a square which indicates the performance level: green when the TEa limit is reached, blue if only the SA limit is met. The number of points is always assigned on the basis of the widest interval. If the TEa limit is met, but only 1 point and for SA 2 points, then this is indicated by an asterisk with foot note explanation. The point shown is then for the (sharper!) TEa limit.

Here too the summary of the difference plot is shown using the same score pictogram as on the summary page of the regular report. In contrast to the regular report all individual points are now processed without time-weighting.

The column Survey scores is intentionally left empty.



Year report 2017

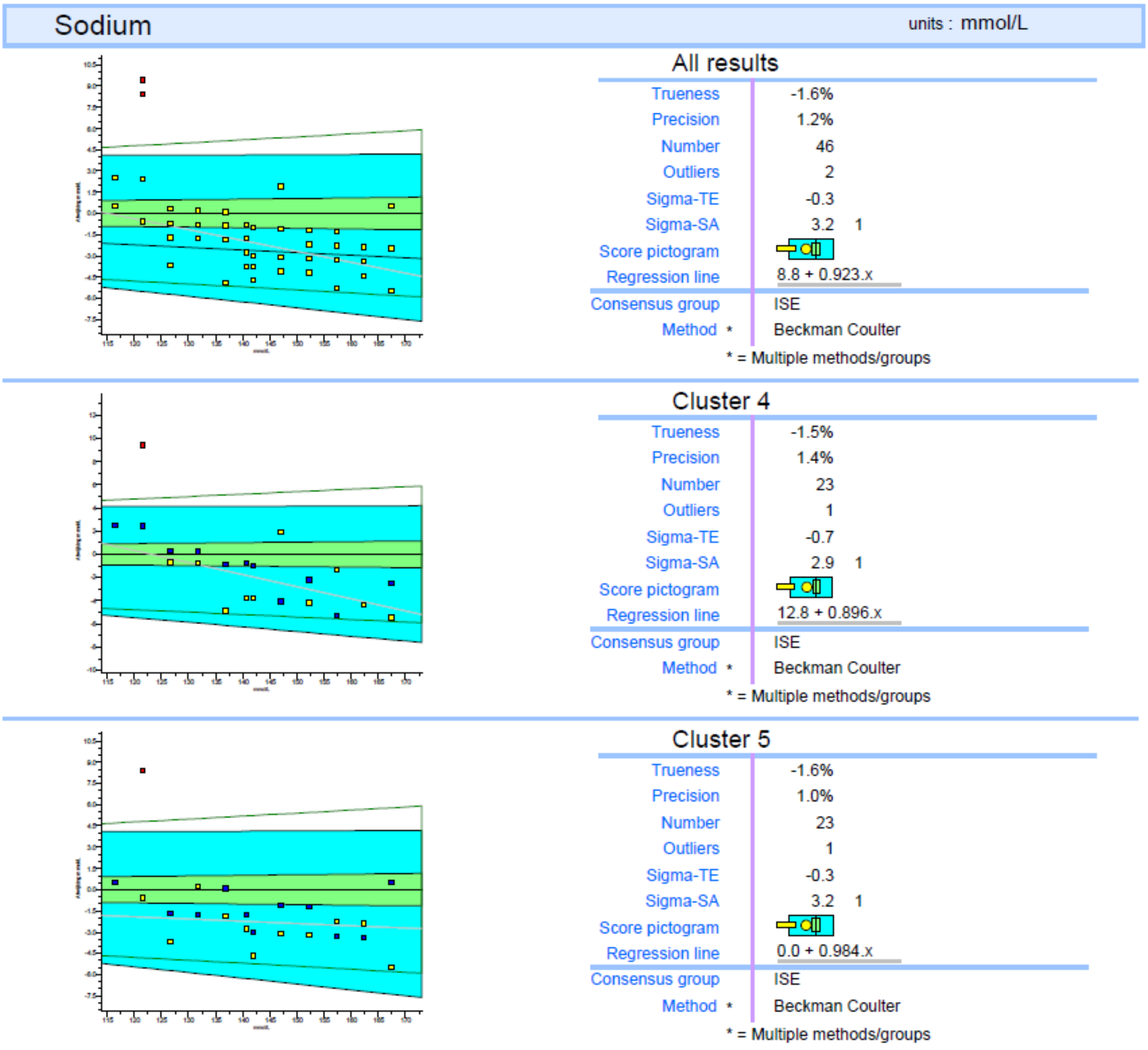
Analyte	Year score	Survey scores
Sodium	2	
Potassium	2	
Chloride	2	
Calcium	1	
Inorg. Phosphate		
Magnesium	1	
Lithium	1	
Iron		
Urea		
Creatinine	1	
Urate		
Glucose	2	
Osmolality		
Lactate		
Total Protein	1	
Albumin		
Bilirubin	2	
Alk. Phosphatase	2	
ASAT	2	
ALAT	2	
LD	1*	
Gamma-GT	2	
CK	2	
Amylase	2	
Lipase		

* = Voor SA 2 punten

Legend 🟩 = Within TE 🟦 = Within SA 🟥 = Outside TE/SA □ = No value 🟩 = Correct 🟥 = Incorrect

In the second part of the report the difference plots are shown per analyte. These difference plots and associated scores are the averaging of all submitted clusters. Thus an image of the between-cluster variation also materializes, which can be helpful with your management review, where after all you also have to assess if the between-analysis variation is controlled adequately. If multiple clusters are submitted, the difference plots per cluster are also shown.

You already have the information per cluster from the regular report, but there the rendition is time-weighted and duplicates are not recognizable as such. Now the duplicates are recognizable as such as a second point of a different colour with the same x-value. By comparing cluster transcending information with information per cluster, you are able to assess to which extent the individual clusters are adapted sufficiently to each other where that is necessary. The cluster transcending information is only shown if the results of all clusters during the whole year are scored against the same reference (either reference value or the same consensus group).



Sodium units : mmol/L

All results

Trueness	-1.6%
Precision	1.2%
Number	46
Outliers	2
Sigma-TE	-0.3
Sigma-SA	3.2 1
Score pictogram	
Regression line	8.8 + 0.923.x
Consensus group	ISE
Method *	Beckman Coulter

* = Multiple methods/groups

Cluster 4

Trueness	-1.5%
Precision	1.4%
Number	23
Outliers	1
Sigma-TE	-0.7
Sigma-SA	2.9 1
Score pictogram	
Regression line	12.8 + 0.896.x
Consensus group	ISE
Method *	Beckman Coulter

* = Multiple methods/groups

Cluster 5

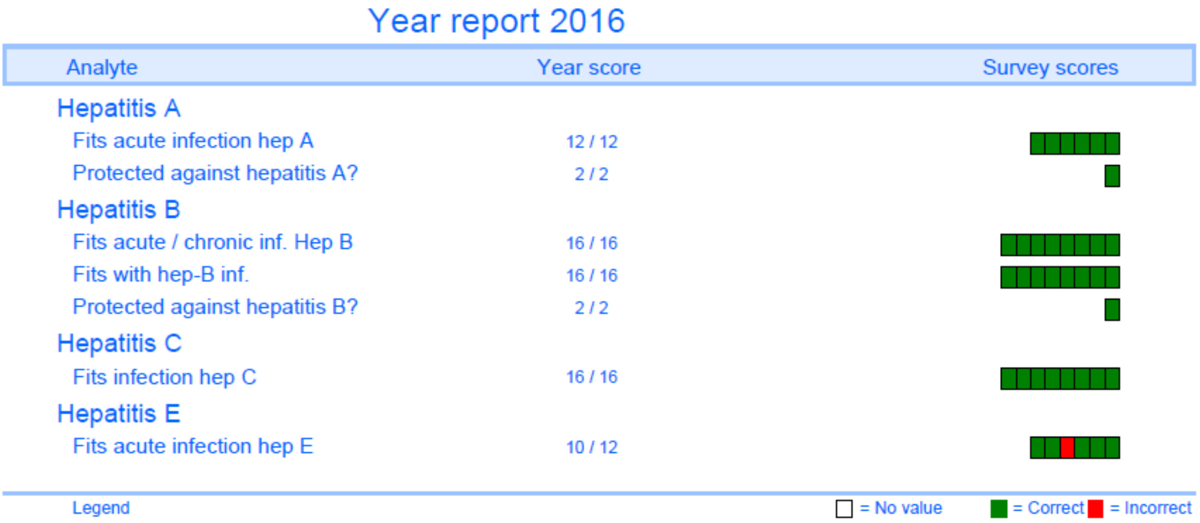
Trueness	-1.6%
Precision	1.0%
Number	23
Outliers	1
Sigma-TE	-0.3
Sigma-SA	3.2 1
Score pictogram	
Regression line	0.0 + 0.984.x
Consensus group	ISE
Method *	Beckman Coulter

* = Multiple methods/groups

6.2. Qualitative annual report

For each analyte, grouped by analyte group, the score indicators are shown. Because not all analytes for each sample are requested, the length of the bars varies per analyte.

In the same report the indicators for the conclusion questions are also shown here. Because the questions asked are "free text" boxes, no further detailing can be provided and all answers are displayed in one bar.



Year report 2016

Analyte	Year score	Survey scores
Hepatitis A		
Fits acute infection hep A	12 / 12	🟩🟩🟩🟩🟩🟩
Protected against hepatitis A?	2 / 2	🟩
Hepatitis B		
Fits acute / chronic inf. Hep B	16 / 16	🟩🟩🟩🟩🟩🟩🟩🟩
Fits with hep-B inf.	16 / 16	🟩🟩🟩🟩🟩🟩🟩🟩
Protected against hepatitis B?	2 / 2	🟩
Hepatitis C		
Fits infection hep C	16 / 16	🟩🟩🟩🟩🟩🟩🟩🟩
Hepatitis E		
Fits acute infection hep E	10 / 12	🟩🟩🟥🟩🟩🟩

Legend □ = No value 🟩 = Correct 🟥 = Incorrect

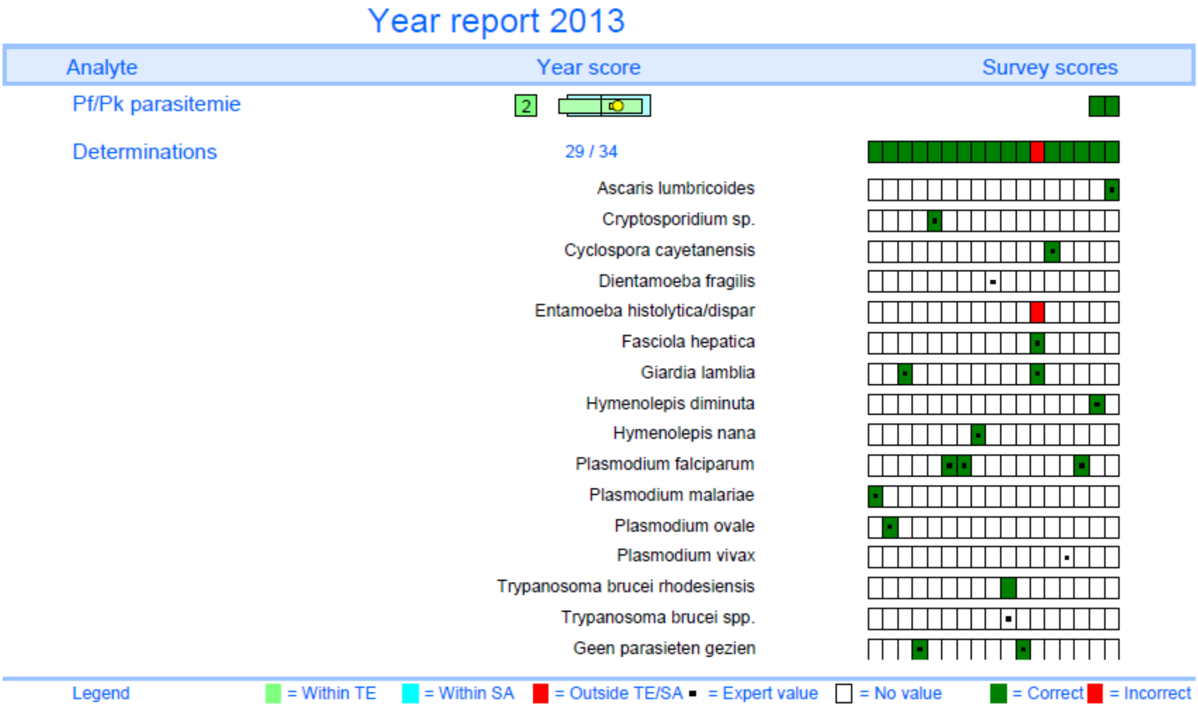
6.3. Conclusion questions annual report

See qualitative annual report above

6.4. Deterministic annual report

In the deterministic annual report the number of correctly determined samples is displayed in the first line. An enumeration of the micro-organisms found in that year is given below that. Each column reflects one sample. Here too a correct determination is indicated by a green colour and an incorrect determination (false positive or false negative) by a red colour (in example below the *Entamoeba histolytica/dispar*).

If the participant has forwarded a sample, although there are functional missed micro-organisms (false negative) this does not lead to an incorrect score of the organism concerned. It is also possible that a participant gives a good, but less precise answer than the expert (eg. *Trypanosoma brucei spp.*, whilst the expert value is *Trypanosoma brucei rhodesiensis*). Here too no incorrect assessment, however the expert value can be read by way of the black dot in the box with the expert result.



Year report 2013

Analyte	Year score	Survey scores
Pf/Pk parasitemie	2	
Determinations	29 / 34	
	Ascaris lumbricoides	
	Cryptosporidium sp.	
	Cyclospora cayetanensis	
	Dientamoeba fragilis	
	Entamoeba histolytica/dispar	
	Fasciola hepatica	
	Giardia lamblia	
	Hymenolepis diminuta	
	Hymenolepis nana	
	Plasmodium falciparum	
	Plasmodium malariae	
	Plasmodium ovale	
	Plasmodium vivax	
	Trypanosoma brucei rhodesiensis	
	Trypanosoma brucei spp.	
	Geen parasieten gezien	

Legend 🟩 = Within TE 🟦 = Within SA 🟥 = Outside TE/SA ■ = Expert value □ = No value 🟩 = Correct 🟥 = Incorrect

●	expert has found determinant, participant has not participated for determinant
	expert has NOT found determinant, participant has not participated for determinant
🟩 ●	expert has found determinant, participant has reported determinant as PRESENT
🟩	expert has NOT found determinant, participant reported determinant as ABSENT
🟥 ●	expert has found determinant, participant reported determinant (wrongly) ABSENT
🟥	expert has NOT found determinant, participant reported determinant (wrongly) PRESENT

7. References

1. Miller WG, Myers WL, Lou Gantzer M, Kahn SE, Schönbrunner ER, Thienpont LM, Bunk DM, Christenson RH, Eckfeldt JH, Lo SF, Nübling CM, and Sturgeon CM. Roadmap for harmonization of clinical laboratory measurement procedures. *Clin Chem* 2011; 57: 1108-1117
2. Miller GW, Jones GRD, Horowitz GL, and Weykamp C. Proficiency testing/external quality assessment: current challenges and future directions. Clin Chem 2011; 57: 1670-1680
3. Miller WG, Myers GL, and Rej R. Why commutability matters. Clin Chem 2006; 52: 553-554
4. Müller MM. Implementation of reference systems in laboratory medicine. Clin Chem 2000; 46: 1907-1909
5. Ricos C, Alvarez V, Cava F, et al. Current databases on biologic variation: pros, cons and progress. Scand J Clin Lab Invest 1999;59: 491– 500.
6. http://www.westgard.com/biodatabase1.htm
7. Jansen RTP. The quest for comparability: Calibration 2000. Accred Qual Assur 2000; 5: 363-366
8. Baadenhuijsen H, Steigstra H, Cobbaert C, Kuypers A, Weykamp C, and Jansen R. Commutability Assessment of Potential Reference Materials Using a Multicenter Split-Patient-Sample Between-Field-Methods (Twin-Study) Design: Study within the Framework of the Dutch Project "Calibration 2000". Clin Chem 2002; 48: 1520-1525
9. Cobbaert C, Weykamp C, Baadenhuijsen H, Kuypers A, Lindemans J, and Jansen R. Selection, Preparation, and Characterization of Commutable Frozen Human Serum Pools as Potential Secondary Reference Materials for Lipid and Apolipoprotein Measurements: Study within the Framework of the Dutch Project "Calibration 2000". Clin Chem 2002; 48: 1526-1538
10. Baadenhuijsen H, Weykamp C, Kuypers A, Franck P, Jansen R, and Cobbaert C. Commuteerbaarheid van het huidige monstermateriaal in de SKML-rondzendingen van de algemene klinische chemie. Ned Tijdschr Klin Chemie Labgeneesk 2008; 33: 154-157
11. Baadenhuijsen H, Kuypers A, Weykamp C, Cobbaert C, and Jansen R. External quality assessment in the Netherlands: time to introduce commutable survey specimens. Lessons from the Dutch "Calibration 2000" project. Clin Chem Lab Med 2005; 43: 304-307
12. Thelen MH, Jansen RT, Weykamp CW, Steigstra H, Meijer R, Cobbaert CM. Expressing analytical performance from multi-sample evaluation in laboratory EQA. Clin Chem Lab Med. 2017 doi: 10.1515/cclm-2016-0970. [Epub ahead of print]
13. Sandberg S, Fraser C, Horvath A, Jansen R, Jones G, Oosterhuis W, 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; 53:833–5.
14. EFLM Biological Variation Database, https://biologicalvariation.eu/meta_calculations