Clinical decision limits as criteria for setting analytical performance specifications for laboratory tests

Rotgers, Emmi

2023-02-01

Clinical decision limits as criteria for setting analytical performance specifications for laboratory tests

Emmi Rotgers [a,1], Solveig Linko [b], Elvar Theodorsson [c], Timo T. Kouri [a,*]

[a] *Department of Clinical Chemistry, University of Helsinki, and HUSLAB, HUS Diagnostic Center, Helsinki University Hospital, FIN-00029 Helsinki, Finland*
[b] *Linko Q-Solutions, FIN-00950 Helsinki, Finland*
[c] *Department of Biomedical and Clinical Sciences, Division of Clinical Chemistry and Pharmacology, Linkoping University, SE-58183 Linkoping, Sweden*

A R T I C L E   I N F O

A B S T R A C T

Background: The biological (CV_I), preanalytical (CV_{PRE}), and analytical variation (CV_A) are inherent to clinical laboratory testing and consequently, interpretation of clinical test results.
Methods: The sum of the CV_I, CV_{PRE}, and CV_A, called diagnostic variation (CV_D), was used to derive clinically acceptable analytical performance specifications (*CAAPS*) for clinical chemistry measurands. The reference change concept was applied to clinically significant differences (*CD*) between two measurements, with the formula $CD = z*\sqrt{2*}CV_D$. *CD* for six measurands were sought from international guidelines. The *CAAPS* were calculated by subtracting variances of CV_I and CV_{PRE} from CV_D. Modified formulae were applied to consider statistical power (*1-β*) and repeated measurements.
Results: The obtained *CAAPS* were 44.9% for urine albumin, 0.6% for plasma sodium, 22.9% for plasma pancreatic amylase, and 8.0% for plasma creatinine ($z = 3$, $α = 2.5\%$, *1-β* = 85%). For blood HbA$_{1c}$ and plasma low-density lipoprotein cholesterol, replicate measurements were necessary to reach *CAAPS* for patient monitoring. The derived *CAAPS* were compared with analytical performance specifications, *APS,* based on biological variation.
Conclusions: The CAAPS models pose a new tool for assessing *APS* in a clinical laboratory. Their usability depends on the relevance of *CD* limits, required statistical power and the feasibility of repeated measurements.

1. Introduction

Laboratories need both clinical and analytical performance specifications to ensure that their measurements are fit for the intended use in patient care. Manufacturers of measuring systems, reagents, and calibrators need these specifications when optimizing and evaluating whether their products are fit for purpose. Governmental regulators around the globe also need them when evaluating IVD devices. The earliest published analytical performance specifications were based on the opinions of "various interested pathologists, other physicians, and

other medical laboratory scientists" [1], on biological variation [2], and on the "state of the art" [3]. A hierarchical classification of the criteria with the effects on clinical outcomes at the pinnacle was proposed [4] and agreed on at the IFCC-IUPAC conference in Stockholm in 1999 [5]. The hierarchical structure was confirmed and simplified at the EFLM strategic conference in Milan in 2014, leaving out opinion criteria [6]. Clinical outcomes (Model 1) and biological variation (Model 2) were preferred strategies depending on the metrological and diagnostic properties of each measurand, leaving the state-of-the-art, Model 3, and combinations of the three models, also as options [6]. Model 1a refers to

direct link between testing and health outcomes and Model 1b refers to impact of laboratory testing on medical decisions and classifications, i. e., an indirect link between testing and health outcomes [7]. The EFLM Task and Finish Group on the allocation of laboratory tests to Different Models, EFLM TFG-DM, has proposed example measurands for the different *APS* models [8]. In practice, setting a universal *APS* even for blood hemoglobin A_{1c} remains challenging after comparing various models [9].

Measurement uncertainty (*MU*), and related analytical performance specifications combine uncertainty of the assigned values of reference materials, uncertainty in the assignment of calibrator values, and imprecision of the reproducibility of results [10–12]. Variation in test results also includes biological and preanalytical variation in addition to the analytical variation [13–15]. It remains to be seen in the future how regulatory and accreditation bodies will assess the clinical performance of laboratory tests in their intended use, against the requirements of the new IVDR regulation [16] and the ISO 15189:2022 standard, Chapter 7.3. [17].

This study was initiated to verify clinical performance of the measurement procedures on a novel automated platform at HUS Diagnostic Center, following the requirements of the ISO 15,189 standard. The described approach was used to model clinically acceptable performance specifications (*CAAPS*) for six measurands, i.e., blood hemoglobin A_{1c} (HbA$_{1c}$), urine albumin (Alb), plasma sodium (Na), plasma pancreatic amylase (AmylP), plasma low-density lipoprotein cholesterol (LDL-C), and plasma creatinine (Crea).

2. Materials and methods

2.1. Criteria for clinically acceptable analytical performance specifications (CAAPS)

Clinical guidelines were used to calculate *CAAPS* by converting them to clinically significant differences (*CD*) for six common clinical chemistry measurands with variable characteristics, as listed in Table 1.

Blood HbA$_{1c}$ is used both for the diagnosis of diabetes mellitus type 2 and for treatment follow-up. The decision limit in the diagnosis of diabetes is 48 mmol/mol or 6.5 Hgb%. In comparison, the upper health-related reference limit is 42 mmol/mol or 6.0 Hgb%, expressed either with IFCC (International Federation of Clinical Chemistry and Laboratory Medicine) reference measurement units, or with NGSP (National Glycohemoglobin Standardization Program) conventional units [18]. In evaluating glycemic control, the general target for treatment is < 53 mmol/mol / <7 Hgb%, or < 64 mmol/mol / <8 Hgb% in cases where less stringent goals are necessary. In the assessment of change, 5 mmol/mol (IFCC unit) or 0.5 Hgb% (NGSP unit) is interpreted as significant both by U.S. and European specialists, indicating a relative change of 9% at 53 mmol/mol or 7% at 7 Hgb% [19,20]. In a survey for general practitioners in six European countries, a decrease in blood HbA$_{1c}$ between 7 and 20% and an increase of 6–10% were deemed relevant when using NGSP units [21]. Furthermore, in the UK Prospective Diabetes Study for the effect of glycemic control on clinical outcomes in patients with type 2 diabetes, the patients were grouped into intensive and standard glycemic control using a difference of 11% in blood HbA$_{1c}$ (NGSP unit) [22]. From these data, we deduced two different levels of clinical need: the significant difference in diagnostic testing is 14%, i.e., (48–42)/42 mmol/mol and 9% corresponding to (6.5–6.0)/6.0 Hgb%, while in monitoring glycemic control it would be about 10% in mmol/mol, and respectively 7% in Hgb%.

Urine albumin (Alb) is used to screen diabetic and hypertensive nephropathy. The recommended assay for albuminuria screening is the urine albumin/creatinine-ratio (ACR), adjusting Alb concentration against volume rate (diuresis). According to the KDIGO guideline, normal albuminuria is below 3 mg/mmol, moderately increased albuminuria is 3–30 mg/mmol, and severely increased albuminuria is>30 mg/mmol [23]. A gradual transition from healthy to nephropathic

Table 1
Clinically significant differences for laboratory analytes.[a,b]

Measurand	Clinically significant difference, CD [c]	CD, in (%)	Chosen clinical setting	Source [a] [Reference used]
Blood HbA$_{1c}$	Diagnostic testing: 6 (42−> 48 mmol/mol) or 0.5 (6.0-> 6.5 Hgb%)	14% (IFCC) or 9% (NGSP)	Diagnosis of diabetes, diagnostic interval in IFCC and NGSP units	WHO Classification of diabetes mellitus 2019 [18]
	Monitoring: 5 mmol/mol or 0.5 Hgb% change	10% (IFCC) or 7% (NGSP)	Worsening of diabetes (most stringent limits) at 53 mmol/ mol (IFCC) or 7.0 Hgb% (NGSP)	Little RR et al 2011 [19], Skeie S et al 2005 [21], Turner RC et al 1998 [22]
Urine Alb	70 (30−> 100 mg/L)	230%	Initial detection of moderate albuminuria (30–300 mg/L), a limit at the logarithmic **midpoint** 100 mg/L of moderate albuminuria range, corresponding to ACR range 3–30 mg/mmol at an average urine Crea concentration of 10 mmol/L	KDIGO Guideline 2012 [23]
Plasma Na	5 (125−> 130 mmol/ L)	4%	Treatment of profound hyponatremia, differentiation of 5 mmol/L	Spasovski G et al, European hyponatraemia guideline 2014 [25]
Plasma AmylP	Upper reference limit (URL -> 2 × URL); e.g., 65 (65−> 130 U/L)	100%	In alerting for a possibility of acute pancreatitis, a cut-off of 2xURL from the **midpoint** of a 4-fold change compared to URL	Tenner S et al, American College of Gastroenterology Guideline 2013 [26]
Plasma LDL-C	0.4 (1.8 -> 1.4 mmol/L; or 0.4 (2.6 -> 2.2 mmol/ L)	20% (22–15%)	Treatment target of dyslipidemia in the very high-risk group for CVD; treatment target for the moderate risk group	Mach F et al, ESC/ EAS Guideline 2020 [27]
Plasma Crea	29 (72−> 101 μmol/L)	40%	Differentiation between limits of normal (eGFR 90 mL/ min/1.73 m^2; P-Crea 72 μmol/ L) and mildly impaired kidney function (eGFR 60 mL/ min/1.73 m^2; P-Crea 101 μmol/ L) in females at 40 years of age [d]	KDIGO Guideline 2012 [23]

a Detailed explanations on selection criteria and their references are given in the text, *Chapter 2.1. Criteria for clinically acceptable analytical performance specifications (CAAPS)*.
b Abbreviations used: ACR, albumin/creatinine ratio; Blood HbA_{1c}, blood hemoglobin A_{1c}; CD, clinically significant difference; CKD-EPI, the Chronic Kidney Disease Epidemiological formula; CVD, cardiovascular disease; EAS, European Atherosclerosis Society; ESC, European Society of Cardiology; eGFR, glomerular filtration rate, estimated; IFCC, International Federation of Clinical Chemistry and Laboratory Medicine, recommended unit for HbA_{1c} measurements; KDIGO, Kidney Disease - Improving Global Outcomes; NGSP, National Glycohemoglobin Standardization Program, conventional unit for HbA_{1c} measurements; Plasma AmylP, plasma pancreatic amylase; Plasma Crea, plasma creatinine; Plasma LDL-C, plasma LDL cholesterol; Plasma Na, plasma sodium; Urine Alb, urine albumin; URL, upper reference limit; WHO, World Health Organization.
c The first value of each interval is the denominator of the relative difference.
d eGFR was estimated by using the CKD-EPI formula.

kidneys is the critical change, requiring differentiation between normal and repeatedly demonstrated moderate albuminuria. Within the exponential interval of 3–30 mg/mmol ACR, an incipient nephropathy must be established, and treatment initiated, already at the logarithmic **midpoint** of the moderate albuminuria range – not at the upper limit of 30 mg/mmol several years later. Thus, a significant difference was chosen to be a change of 3 → 10 mg/mmol, corresponding to a change of 30 to 100 mg/L urine Alb divided by an average urine creatinine concentration of 10 mmol/L. This prognostically significant detection of kidney disease corresponds to a difference of 230%. Adjusting urine albumin concentration to that of urine creatinine is important because it reduces the CV_I of ACR to about 30% in random spot specimens, corresponding to the CV_I of Alb concentration in first-morning urine [24].

Plasma sodium (Na) is used to assess water and electrolyte balance. Hyponatremia is classified as mild at 130–135 mmol/L, moderate at 125–129 mmol/L, and profound at<125 mmol/L. During the correction of hyponatremia with intravenous infusion of hypertonic NaCl solution, the treatment goal for plasma Na increase is 5 mmol/L per 24 h, and it should not exceed 8 mmol/L per 24 h prior to reaching 130 mmol/L [25]. The goal of detecting a 5 mmol/L difference in plasma Na concentration is clearly desirable, indicating a significant difference of 5/ 125 = 4% at concentrations < 130 mmol/L. Detecting an increase of 8 mmol/L was an absolute minimum specification. In the treatment of hypernatremia, it is significant to detect a decrease of 10 mmol/L in 24 h using repeated measurements, yielding a significant difference of 7% for Na concentrations above the upper reference limit (URL).

Plasma pancreatic amylase (AmylP) is used to diagnose acute pancreatitis in patients with acute abdominal pain. AmylP activity>3–5 times URL (65 U/L at HUS Diagnostic Center) is diagnostic for acute pancreatitis if observed together with abdominal pain or imaging that is consistent with pancreatitis according to the American College of Gastroenterology Guideline [26]. The average significant difference for plasma AmylP activity for diagnosis of acute pancreatitis would be 300% to reach a high diagnostic specificity (Sp) of the result (1xURL –> 4xURL; Sp > 95%). In clinical practice, plasma AmylP activity has a reduced sensitivity (Sn) in detection of alcoholic pancreatitis or severe necrotic pancreatitis, and in patients with symptoms that have lasted for several days (Sn < 50% after 4 days of disease onset). Because of these risks, a lower decision limit was modelled to alert for the possibility of pancreatitis, using a significant difference of 2xURL or 100% increase in plasma AmylP activity (65 to 130 U/L).

Plasma low-density lipoprotein cholesterol (LDL-C) is used to assess hyperlipidemia, a risk factor for cardiovascular disease (CVD). In the recent 2019 ESC/EAS Guidelines for the Management of Dyslipidemias, treatment targets for plasma LDL-C were redefined into concentrations of 1.4, 1.8, 2.6, and 3.0 mmol/L depending on the level of CVD risk [27]. The feasibility of differentiating the several categories require the detection of differences of about 0.4 mmol/L in plasma LDL-C concentrations. A clinically significant difference is then 22% between the

highest risk limits (1.8 mmol/L and 1.4 mmol/L) or 15% starting from the upper limit of the moderate risk individuals (0.4/2.6 mmol/L = 15%). An average estimate of 20% was used in further calculations.

Plasma creatinine (Crea) is used mainly to estimate glomerular filtration rate (eGFR) for the evaluation of kidney function. Calculated from the CKD-EPI formula for eGFR, the limits of a mildly impaired kidney function (from 90 to 60 mL/min/1.73 m^2) were used to estimate the clinically significant difference, below which active diagnostics and treatment are indicated, since eGFR < 60 mL/min/1.73 m^2 suggests moderately impaired kidney function [23]. These were modelled in 40-year-old women with the smallest changes in plasma Crea needed to differentiate the given GFR estimates. A corresponding change in plasma Crea concentration of 40% (from 72 to 101 μmol/L) was used.

2.2. Calculations of CAAPS

2.2.1. Calculating diagnostic variation, CV_D

Results of quantitative laboratory measurements represent point estimates with uncertainty distributions around the obtained value. Clinical laboratory tests are typically used to compare two consecutive results (monitoring of a patient), or to differentiate diseases from a healthy state or to establish prognostic categories (diagnostic testing). Then, the combined variation, defined as *diagnostic variation, CV_D*, includes analytical variation, CV_A, preanalytical variation, CV_{PRE}, and intra-individual biological variation, CV_I [15,28]. Two overlapping Gaussian distributions were used to model clinically significant differences, *CD* [29]. The *CD* between two measurement values was used to calculate the clinically acceptable CV_D of a laboratory result using the conventional formula of relative change value, *RCV* [21,28]:

$$CD = z* \sqrt{2*CV_D} \qquad (1)$$

where z is the Gaussian statistic, CV_D = diagnostic variation, and $\sqrt{2}$ assumes two identical uncertainty distributions in the compared measurements. By converting the formula (1), the acceptable CV_D was obtained as follows:

$$CV_D = CD/(z*\sqrt{2}) \qquad (2)$$

2.2.2. Sources of variation in the pre-examination processes

Data for biological intra-individual variation, CV_I, were mostly obtained from the database provided by the Task and Finish Group of the European Federation of Clinical Chemistry and Laboratory Medicine, EFLM [30]. The provided median was used as the best estimate of intra-individual biological variation, CV_I, as reported by the EFLM Working Group on Biological Variation.

Estimates for preanalytical technical variation are usually ignored, since standardized sample collection and sample handling procedures are assumed to minimize it [28,31]. However, we included this factor to allow estimation of possible effects of regional storage and transportation on sample quality in practical assessments of acceptability of samples [32,33]. An acceptable preanalytical variation, CV_{PRE}, was estimated based on our previous study [14], testing the effect of regional transportation: a 1% variation was estimated for high-concentration, stable measurands, such as blood HbA_{1c}, plasma LDL-C and plasma Crea. Four per cent was allowed for plasma AmylP activity due to a possible inactivation during storage, and 5% for urine Alb, due to its tendency to adhere on the walls of specimen containers. For plasma Na, 0.5% preanalytical variation was used.

2.2.3. Determining acceptable analytical variation CV_A from CV_D

The square of combined diagnostic variation, CV_D^2, was summarized from its variance components [34], as follows:

$$CV_D^2 = CV_I^2 + CV_{PRE}^2 + CV_A^2 \qquad (3)$$

From this equation, the acceptable analytical variation, CV_A was left

over after subtracting variances of the other components from the variance of CV_D (Fig. 1), adopting an earlier example on postanalytical quality assessment [21]:

$$CV_A^2 \leq CV_D^2 - \left(CV_I^2 + CV_{PRE}^2\right) \tag{4}$$

The resulting clinically acceptable analytical variation, now called clinically acceptable analytical performance specification, $CAAPS = \sqrt{CV_A^2}$, is then required to detect the clinically significant difference CD used in the modeling.

2.2.4. Statistical power and budget for diagnostic variation, CV_D

A difference that is considered clinically significant must be detected with acceptable sensitivity and specificity to justify a decision on a clinical measure (further investigation or treatment). In addition to providing a statistical estimate of still a stable situation (false positives, probability α of type I error), as applied in the classical RCV [28,29], another estimate, the alternative probability β of type II error (false negatives) was used. It describes the sensitivity, or statistical power of an observed difference, $1-\beta$ [35]. Detection of a CD between two measurement values depends on the statistical sensitivity, resulting in different uncertainty budgets for CV_D according to the used z statistic in the equation (2). The CV_D budgets were first calculated with the commonly used borderline statistical sensitivity $1-\beta = 50\%$ (at $z = 1.96$). To reach a sufficient statistical power to detect a clinically significant change, the z statistic needs to be increased from the conventional 1.96 (bidirectional change at $2\alpha = 5\%$; or $\alpha = +/- 2.5\%$ for a false positive detection), since at $z = 1.96$ a sensitivity of just $1-\beta = 50\%$ is obtained (Fig. 2). A change of $z = +3$ has a sensitivity of 85%, and a change up to $z = +4$ a sensitivity of 98% in detecting a unidirectional change, when keeping the unidirectional $\alpha = 2.5\%$ [29]. Originally, statistical power functions were developed by James O. Westgard and coworkers for error detection in statistical process control in 1970's [36].

Fig. 1. The flowchart on estimation of clinically acceptable analytical performance specification (CAAPS) from clinically significant difference (CD). Step A: Two results represent a CD for a measurand in a defined clinical situation. Step B: The budget of diagnostic variation, CV_D, is obtained by dividing CD with $z*\sqrt{2}$, where z is the chosen Gaussian statistic. Step C: Biological intra-individual variation (CV_I) and preanalytical technical variation (CV_{PRE}), are subtracted from CV_D as squared terms (variances) to obtain the variation left for analytical performance (CV_A) that represents the CAAPS.

2.2.5. Need for repeated measurements

When it is impossible to detect a change between two measured values with sufficient statistical power due to a large biological variation or analytical variation, CV_D may be achieved using n replicate samples. This reduces the variation by division with a factor of $\sqrt{n}$ [37]. The acceptable diagnostic variation from repeated measurements, CV_{REP}, is calculated as follows:

$$CV_D = CV_{REP}/\sqrt{n}, \text{ where } n = \text{number of repeats} \tag{5}$$

Consequently, clinically significant difference $CD = z * \sqrt{2} * CV_D = z * \sqrt{2} * CV_{REP} / \sqrt{n}$.
Inversely, an acceptable $CV_{REP} = CV_D * \sqrt{n}$.

3. Results

Prognostic (urine Alb; plasma LDL-C), or diagnostic groups (plasma AmylP; plasma Crea) we associated with wide CD estimates using either limits or midpoint values of the neighboring categories, while measurands used mostly for monitoring (blood HbA$_{1c}$; plasma Na) showed narrow CD estimates (Table 1).

The *CAAPS* estimates were initially calculated with the most used RCV at $p < 5\%$ (2α, $z = 1.96$), using equations (2)–(4) (Table 2). For plasma LDL-C, the intra-individual variation CV_I was already larger than the obtained allowable CV_D calculated with the borderline statistical power $1-\beta = 50\%$ (Fig. 2). For blood HbA$_{1c}$, the obtained *CAAPS* estimates were notably narrow as well, and even narrower for NGPS units than for IFCC units in both diagnostic testing and monitoring. The calculated *CAAPS* was very stringent for plasma sodium as well (Table 2).

To improve the statistical power, we calculated acceptable *CAAPS* estimates at sensitivities of 85% ($z = +3$), and 98% ($z = +4$) at unidirectional $\alpha = 2.5\%$ [29] (Table 3). When increasing the z statistic to $+4$ and keeping a given CD, the respective budget for CV_D was decreased according to equation (2). The impact of increasing $1-\beta$ on the calculated CV_D is shown in Table 3. The corresponding *CAAPS* tightened even more than the CV_D suggested, because biological and preanalytical technical variations remained the same. Diagnostic detection of diabetes with HbA$_{1c}$ using a single measurement seemed to be possible at a sensitivity of 85% ($z = +3$) with IFCC unit reporting with a performance of *CAAPS* 1.9%, but not for reporting with NGSP units (*CAAPS* 0.8%).

CAAPS of urine Alb, plasma AmylP and plasma Crea seemed attainable even at a 98% sensitivity ($z = +4$), because of their large CD estimates from diagnostic use (Table 3). On the other hand, the calculated *CAAPS* was tight for plasma Na already at $z = 1.96$. For plasma LDL-C, detection of the ascribed CD was not at all possible, as shown already in Table 2. Using $z = 1.64$ with $p < 10\%$ for false positives, instead of $z = 1.96$, would improve the sensitivity $1-\beta$ from 50% to 64%, but still remain insensitive in detecting clinically significant changes.

Repeating a laboratory measurement using a new specimen is commonly applied to confirm the detection of a change. To model this, equation (5) was used. The effect of repeated sampling and measurements on CV_D and subsequent *CAAPS* was calculated by multiplying CV_D with $\sqrt{n}$, to obtain allowable CV_{REP}, when using $n = 1$ to 4 (Table 4). With a statistical power $1-\beta = 85\%$ ($z = 3$) and a unidirectional $\alpha = 2.5\%$, a reasonable *CAAPS* was reached using duplicate measurements for monitoring blood HbA$_{1c}$ (IFCC units), triplicates for plasma Na and monitoring of blood HbA$_{1c}$ (NGSP units), and four replicate measurements to assess a change in plasma LDL-C (Table 4). Detailed calculations of Table 4 are shown for clarity in the Supplementary Table.

The obtained *CAAPS* were compared with *APS* for allowable *MU,* expressed as allowable analytical variation after elimination of bias with calibration, desirably $CV_A \leq 0.5*CV_I$ [38] (Table 5). Also, the conventional *APS* expressed as total allowable error, *TE,* were calculated, with separate bias and imprecision estimates from biological variation [4], despite becoming easily too wide [30]. A comparison of the *CAAPS* and
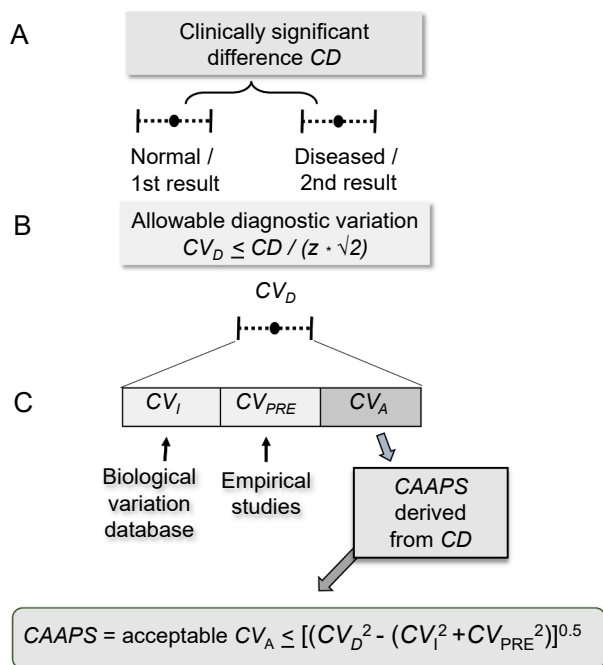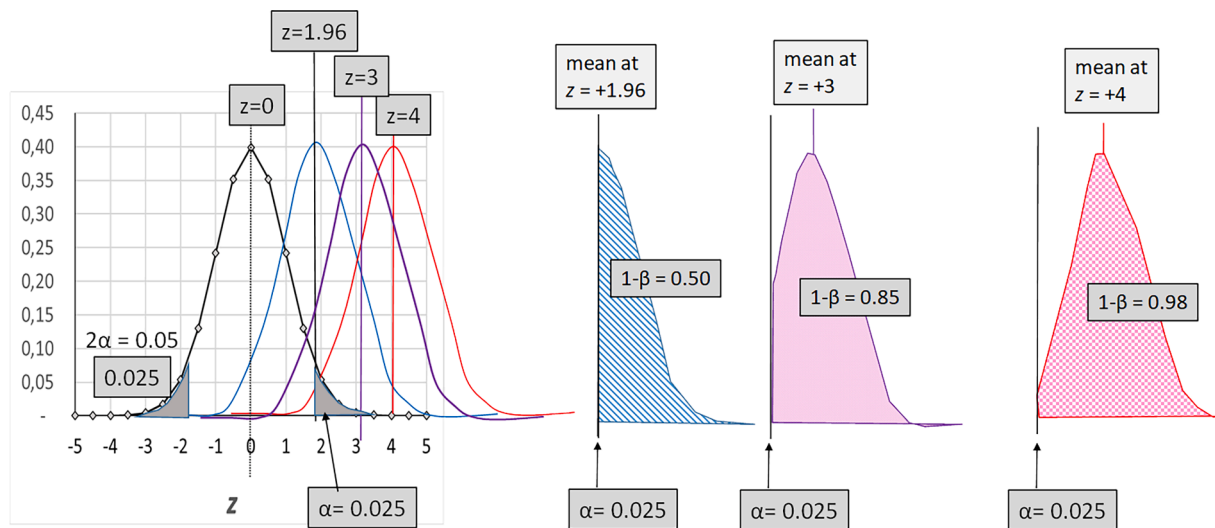
Fig. 2. Effect of size of difference between two results on sensitivity of detection. Statistical power (*1-β*) = sensitivity of detection increases from 0.50 to 0.98, when the difference between measured means increases from $z = +1.96$ to $z = +4$ of Gaussian distribution, using a fixed probability of false positives $α = 0.025$ (shown with a line at $z = +1.96$ of original distribution). Modified from N Iglesias Canadell et al. Clin. Chem. Lab. Med. 42 (2004) 415–422 [29].

Table 2
Clinically acceptable analytical performance specification from clinically significant difference (at $z = 1.96$).

Measurand	Significant difference for medical decision	Clinically significant difference, %	Clinically acceptable diagnostic variation [a]	Former squared	Biological intra-individual variation [b] %	Former squared	Preanalytical variation, estimated	Former squared	Variance remaining for analytical variation [a]	CAAPS[c] based on clinical difference, %
		CD [c]	$CV_D = CD / (1.96*\sqrt{2})$	CV_D^2	CV_I	CV_I^2	CV_{PRE}	CV_{PRE}^2	$CV_A^2 = CV_D^2 - CV_I^2 - CV_{PRE}^2$	$CAAPS = \sqrt{CV_A^2}$ $α = 2.5\%, 1-β = 50\%$
Blood HbA$_{1c}$	Diagnostic testing: 42 -> 48 mmol/mol (IFCC)	14%	5.1%	0.00255	2.5%	0.00063	1%	0.00010	0.00183	4.3% (IFCC)
	6.0 ->6.5 Hgb % (NGSP)	9%	3.2%	0.00106	1.7% [41]	0.00029	1%	0.00010	0.00067	2.6% (NGSP)
Blood HbA$_{1c}$	Monitoring:									
	at 53 mmol/mol (IFCC)	10%	3.6%	0.00130	2.5%	0.00063	1%	0.00010	0.00058	2.4% (IFCC)
	at 7.0 Hgb% (NGSP)	7%	2.5%	0.00064	1.7% [41]	0.00029	1%	0.00010	0.00025	1.6% (NGSP)
Urine Alb	30 -> 100 mg/L	230%	83%	0.68944	30% [24]	0.09000	5%	0.00250	0.59694	77%
Plasma Na [a]	125 -> 130 mmol/L	4%	1.4%	0.00021	0.5% [30]	0.00003	0.5%	0.00003	0.00016	1.3%
Plasma AmylP	URL -> 2 × URL	100%	36%	0.13033	4.0% [30]	0.00160	4%	0.00160	0.12713	36%
Plasma LDL-C [a]	1.8 -> 1.4 mmol/L	20%	**7.2%**	0.00521	**8.0%** [30]	0.00640	1%	0.00010	−0.00129	(<0%)
Plasma Crea	72 -> 101 µmol/L	40%	14.4%	0.02085	4.9% [30]	0.00240	1%	0.00010	0.01835	13.5%

[a] The allowable diagnostic variation CV_D was calculated using the formula of reference change value for a Gaussian distribution: $CV_D = CD / (z*\sqrt{2})$, equation (1). The variance remaining for analytical variation was calculated with equation (4). For plasma LDL-C, detection of a change between two individual measurements was not possible at $z = 1.96$, due to a high CV_I (marked **bold**). Replicate testing is shown in Table 4.

[b] References used for estimates of intra-individual biological variation were the following: [41] Biological variation of diabetics, given NGSP units converted also to IFCC units from S. Carlsen, et al, Clin. Chem. Lab. Med. 2011; [24] S.S. Waikar et al, Am. J. Kidney Dis. 2018; and [30] A.K. Aarsand, et al, EFLM Biological Variation Database, 2022.

[c] Abbreviations used: Blood HbA$_{1c}$, blood hemoglobin A$_{1c}$; *CAAPS*, clinically acceptable analytical performance specification; *CD*, clinically significant difference; CV_A = (acceptable) analytical variation; CV_D = (total) diagnostic variation; CV_I, biological intra-individual variation; CV_{PRE}, preanalytical (technical) variation; IFCC, International Federation of Clinical Chemistry and Laboratory Medicine, mmol/mol unit; NGSP, National Glycohemoglobin Standardization Program, Hgb% unit; Plasma AmylP, plasma pancreatic amylase; Plasma Crea, plasma creatinine; Plasma LDL-C, plasma LDL cholesterol; Plasma Na, plasma sodium; Urine Alb, urine albumin; *z*, Gaussian statistic; *α*, type I error in statistical testing (false positives); *β*, type II error (false negatives); *1-β*, statistical power, sensitivity to detect a change (opposite probability to *β*).

Table 3
Impact of statistical power (1-β) on clinical performance specification.

Z score and associated statistical power $(1-\beta)$ [a]		Mean at (z = 1.96), $(1-\beta)$ = 50%		Mean at (z = 3), $(1-\beta)$ = 85%		Mean at (z = 4), $(1-\beta)$ = 98%	
Measurand	Clinically significant difference, CD %	Acceptable CV_D [b]	CAAPS [c]	Acceptable CV_D [b]	CAAPS [c]	Acceptable CV_D [b]	CAAPS [c,d]
Blood HbA$_{1c}$ (IFCC, DT)	14%	5.1%	4.3%	3.3%	**1.9%**	2.5%	(<0%)
Blood HbA$_{1c}$ (NGSP, DT)	9%	3.2%	**2.6%**	2.1%	0.8%	1.6%	(<0%)
Blood HbA$_{1c}$ (IFCC, Mon)	10%	3.6%	**2.4%**	2.4%	(<0%)	1.8%	(<0%)
Blood HbA$_{1c}$ (NGSP, Mon)	7%	2.5%	1.6%	1.6%	(<0%)	1.2%	(<0%)
Urine Alb	230%	83.0%	77.2%	54.2%	44.9%	40.7%	**27.0%**
Plasma Na	4%	1.4%	**1.3%**	0.9%	0.6%	0.7%	0.0%
Plasma AmylP	100%	36.1%	35.6%	23.6%	22.9%	17.7%	**16.7%**
Plasma LDL-C	20%	7.2%	(<0%)	4.7%	(<0%)	3.5%	(<0%)
Plasma Crea	40%	14.4%	13.5%	9.4%	8.0%	7.1%	**5.0%**

[a] The statistical power $(1-\beta)$ of testing was taken at the unidirectional Gaussian probability of false positives α = +2.5 % (z = 1.96), while increasing the difference of the mean of changed values (Fig. 2), according to N. Iglesias Canadell, P. Hyltoft Petersen, E. Jensen, C. Ricós, E. Jørgensen, Reference change values and power functions, Clin. Chem. Lab. Med. 42 (2004) 415–422.[29].

[b] The diagnostic variations CV_D at different z scores were calculated with the equation (2): $CV_D = CD / (z * \sqrt{2})$.

[c] CAAPS were calculated as shown in Table 2. Limits of achievable CAAPS ranges based on our experience are marked **bold**.

[d] Abbreviations used: Blood HbA$_{1c}$, blood hemoglobin A$_{1c}$; CAAPS, clinically acceptable analytical performance specification; CD, clinically significant difference; CV_D, diagnostic variation; DT, diagnostic testing; IFCC, International Federation of Clinical Chemistry and Laboratory Medicine, mmol/mol unit; Mon, monitoring; NGSP, National Glycohemoglobin Standardization Program, Hgb% unit; Plasma AmylP, plasma pancreatic amylase; Plasma Crea, plasma creatinine; Plasma LDL-C, plasma LDL cholesterol; Plasma Na, plasma sodium; Urine Alb, urine albumin; z, Gaussian statistic; α, type I error (false positives); β, type II error (false negatives); $1-\beta$, statistical power, sensitivity to detect a change (opposite probability to β).

Table 4
Impact of repeated measurements on CAAPS.[b]

Measurand	Clinically significant difference, CD %	Acceptable CV_{REP} after n repeated specimens to reach the required CV_D [a]				CAAPS [b,c] assuming variable number of repeats, n (1 to 4) using z = 3				Biological intra-individual variation, %
Number of repeats (n)		1	2	3	4	1	2	3	4	CV_I
Blood HbA$_{1c}$ (IFCC, Monitoring)	10%	2.4%	3.3%	4.1%	4.7%	(<0%)	**2.0%**	3.1%	3.9%	2.5%
Blood HbA$_{1c}$ (NGSP, Monitoring)	7%	1.6%	2.3%	2.9%	3.3%	(<0%)	1.2%	**2.1%**	2.6%	1.7%
Urine Alb	230%	54.2%	76.7%	93.9%	108.4%	**44.9%**	70.4%	88.8%	104.1%	30.0%
Plasma Na	4%	0.9%	1.3%	1.6%	1.9%	0.6%	1.1%	**1.5%**	1.7%	0.5%
Plasma AmylP	100%	23.6%	33.3%	40.8%	47.1%	**22.9%**	32.8%	40.4%	46.8%	4.0%
Plasma LDL-C	20%	4.7%	6.7%	8.2%	9.4%	(<0%)	(<0%)	1.3%	**4.9%**	8.0%
Plasma Crea	40%	9.4%	13.3%	16.3%	18.9%	**8.0%**	12.4%	15.5%	18.2%	4.9%

[a] Acceptable diagnostic variation of repeated measurements, CV_{REP}, was calculated by using the equation (3), $CV_{REP} = CV_D * \sqrt{n}$. The following derivation of equation (1) applies: $CD = z * \sqrt{2} * CV_D = z * \sqrt{2} * CV_{REP} / \sqrt{n}$. The statistical power $(1-\beta)$ = 85 % with a mean value at z = 3 was used to define the clinically significant change at a unidirectional α = 2.5% (Fig. 2).

[b] Abbreviations used: Blood HbA$_{1c}$, blood hemoglobin A$_{1c}$; CAAPS, clinically acceptable analytical performance specification; CD, clinically significant difference; CV_A, analytical variation; CV_D, diagnostic variation; CV_I, intra-individual biological variation; CV_{REP}, variation of repeated measurements; IFCC, International Federation of Clinical Chemistry and Laboratory Medicine, mmol/mol unit; Mon, monitoring; NGSP, National Glycohemoglobin Standardization Program, Hgb% unit; n, number of repeats; Plasma AmylP, plasma pancreatic amylase; Plasma Crea, plasma creatinine; Plasma LDL-C, plasma LDL cholesterol; Plasma Na, plasma sodium; Urine Alb, urine albumin; z, Gaussian statistic, $1-\beta$, statistical power, sensitivity to detect a change.

[c] CAAPS (acceptable CV_A) were calculated using the equation $CV_A^2 = CV_{REP}^2/n - CV_I^2 - CV_{PRE}^2$ (see also Table 2). Achievable CAAPS ranges based on our experience are shown **bold**.

other estimates of APS for the studied measurands was performed at desirable levels of the other estimates. Compared APS and TE were based on biological variation (Milan Model 2) derived from healthy individuals, except the alternative APS for HbA$_{1c}$ (CV_I of diabetes patients) and another for urine Alb (CV_I of albuminuria patients). Another Milan Model 1b example was listed based on classification errors for blood HbA$_{1c}$ [39]. APS were obtained from published CV_I estimates [24,30,40,41]. For urine Alb, we used additionally a desirable bias of +/-10% against isotope dilution mass spectrometry (ID-MS), with a maximum analytical imprecision of 8% to calculate TE = 23.2% [42]. For plasma Na, consolidated recommendations of TE from EQA schemes were also listed for comparison [43] (Table 5). CAAPS derived from detection of pathophysiological states were generally wider than APS based on CV_I in health.

4. Discussion

4.1. Applicability of the obtained CAAPS

4.1.1. Blood HbA$_{1c}$

A CAAPS for monitoring of blood HbA$_{1c}$ (2.4% in IFCC units; 1.6% in NGSP units) was more stringent than that of diagnostic testing (4.3% IFCC; 2.6% NGSP) (Table 2). A single platform is usually used to perform the HbA$_{1c}$ assay in the laboratories meaning the CAAPS for monitoring would be applied. The CAAPS for monitoring was no longer achievable with the generally used CD when the statistical power $1-\beta$ was increased to 85%, challenging also diagnostic testing with a CAAPS of 1.9% (IFCC), or 0.8% (NGSP) (Table 3). According to this modeling, repeated follow-up samples (Table 4) would be required in monitoring of glycemic control, considering the current performance for blood HbA$_{1c}$ assays [39].

Table 5

Comparison of clinically acceptable analytical performance specifications, *CAAPS,* to other estimates of total allowable error, *TE,* or *APS*.[a]

Measurand	*CAAPS* [b,] singleton comparison at $z = 3$ (Table 3), %	*CAAPS* with replicates (*n*) (Table 4), %	*TE* [c] from literature, %	*APS* from biological variation [d], %	Source [e]	Notes
Blood HbA$_{1C}$	*Diagnostic testing:*		3.0 (IFCC)		[39,40]	The model used $CV_A \leq 3\%$ (IFCC) with no bias causing 2% misclassifications
	1.9 (IFCC)		3.1 (IFCC)	0.8 (IFCC)	[30]	Using CV_I of healthy individuals
	0.8 (NGSP)		2.2 (NGSP)	0.6 (NGSP)	[30]	
	Monitoring:					
	(< 0)	2.0 (*n* = 2; IFCC)	3.9 (IFCC)	1.25 (IFCC)	[41]	Using CV_I of diabetics
	(< 0)	2.1 (*n* = 3; NGSP)	2.7 (NGSP)	0.85 (NGSP)	[41]	
Urine Alb	44.9		23.2	15.0	[24,42]	Using CV_I of albuminuria patients $CV_A \leq 8\%$ (11/17 procedures), $B \leq 10\%$ (8/17 procedures at 30 mg/L albumin); calculated from these, $TE = 23.2\%$
Plasma Na	0.6	1.5 (*n* = 3)	0.7	0.25	[30,40]	
			2.9		[43]	EQA recommendation: +/-4 mmol/L (4 mmol/140 mmol = 2.9 %)
Plasma AmylP	22.9		12.2	2.0	[30]	
Plasma LDL-C	(< 0)	4.9 (*n* = 4)	13.7	4.0	[30]	
Plasma Crea	8.0		7.4	2.25	[30]	

[a] Expressed as percentage (%) for all estimates.

[b] Abbreviations: *APS*, analytical performance specification; *B*, bias; *BV*, biological variation; *CAAPS*, clinically acceptable analytical performance specification; CV_A, analytical variation; CV_G, between-subject biological variation (healthy individuals); CV_I, intra-individual biological variation; EQA, external quality assessment; IFCC, International Federation of Clinical Chemistry and Laboratory Medicine, mmol/mol unit; NGSP, National Glycohemoglobin Standardization Program, Hgb% unit; *TE*, total (allowable) error.

[c] Total allowable error $TE \leq [1.65 \times 0.5\ CV_I + 0.25\ (CV_I^2 + CV_G^2)^{0.5}]$, calculated from biological variation.

[d] APS for random variation of measurements, desirable $CV_A = 0.5 * CV_I$, assuming no bias after calibrations [41].

[e] References in this table were the following (details given in the list of References): [24] S.S. Waikar et al, Am. J. Kidney Dis. 2018;[30] A.K. Aarsand, et al, EFLM Biological Variation Database, desirable APS shown (https://biologicalvariation.eu, accessed 2022);[39] A.A. Nielsen, et al, Clin Chem Lab Med 2014;[40] F. Braga, M. Panteghini, Clin Chem Lab Med 2021;[41] S. Carlsen, et al, Clin Chem Lab Med 2011; [42] L.M. Bachmann et al, Clin Chem 2014; [43] S. Westgard, (https://www.westgard.com/consolidated-goals-chemistry.htm, accessed 2022);

4.1.2. Urine albumin, plasma creatinine and plasma pancreatic amylase

CAAPS were calculated for urine Alb and plasma Crea due to their key role in diagnosing chronic kidney diseases [23]. The diagnostic classification of acute kidney injuries (AKI) also utilizes measurements of plasma creatinine, but well-known limitations of its diagnostic performance excluded AKI from our considerations [44]. The clinically required differences between health and nephropathy-related values could be detected with two singleton results, using *1-β* = 85% (*z* = 3) with an estimated *CAAPS* of 44.9% (urine Alb) and 8.0% (plasma Crea), or even an *CAAPS* = 27.0% and 5.0% with *1-β* = 98% (*z* = 4), respectively (Table 3). Similarly, measurements of plasma AmylP were estimated to allow a *CAAPS* of 22.9% (*1-β* = 85%), or 16.7% (*1-β* = 98%) for diagnostic classification of suspected pancreatitis in emergency room (Table 3). For these measurands, the estimated *CAAPS* were applicable in the chosen clinical settings.

4.1.3. Plasma low-density lipoprotein cholesterol

The large biological intra-individual variation of plasma LDL-C concentrations limits the detection of clinically significant treatment effects in a singleton comparison at a sufficient sensitivity (Table 2). To confirm a 20 % change from 1.8 to 1.4 mmol/l after statin treatment, four repeated measurements were initially needed, and another four measurements in the follow-up according to the Gaussian model at *1-β* = 85% (Table 4). The CVD risk assessment would benefit from another measurand with a smaller CV_I. Unfortunately, the median CV_I of apolipoprotein B is also 7.4% [30]. Larger differences, such as detection of familial hypercholesterolemia with a difference between 5 and 3 mmol/L in plasma LDL-C, are attainable from two singleton results at a sensitivity of 85% with the presented *CAAPS* modeling (data not shown). This is compatible with the ESC/EAS guideline recommendation to detect a −50% reduction of plasma LDL-C concentration as compared to the initial value in high-risk patients [27]. Use of plasma LDL-C in risk assessment tolerates the repetitions as opposed to, e.g., tumor markers

that are used for rapid therapeutic decisions of cancer patients [45]. Statistical sensitivity (*1-β*), urgency of detection from a single specimen, and a possibility to limit the clinical need to a larger diagnostic difference must be considered, when measurands appear to show too wide biological variation.

4.1.4. Plasma sodium

CAAPS that allows detection of an analytical or clinical change of 5 mmol/L in plasma Na was studied, despite the small relative difference. As an example, an emergency patient may have a true plasma Na 124 mmol/L, but it is reported as 129 mmol/L, using the difference of 5 mmol/L corresponding to the limiting *CD* of + 4% (Table 2). A false increase could be excluded with the calculated *CAAPS* of 1.3% (*z* = 1.96, α = 2.5%). However, if a statistical power of *1-β* = 85% is expected in detection of a change in analytics, the acceptable *CAAPS* diminished to as low as 0.6% with singleton measurements (Table 3). After triplicate measurements both originally and in the follow-up, a currently attainable *CAAPS* = 1.5% was reached at *1-β* = 85% (Table 4). Both repeated sampling and transportation delays to central laboratory indicate that point-of-care devices are to be adopted in plasma sodium diagnostics in intensive care.

CAAPS frames seem to provide generally useful estimates of acceptable overall analytical variation for clinical laboratory measurements. In addition, clinical laboratories need narrower limits to internal quality control rules of their analytical measurements to be able to guarantee day-to-day reproducibility in their clinical laboratory environments.

4.2. Comparison of CAAPS with other APS estimates

CAAPS derived from combined diagnostic variation, CV_D (Fig. 1), represent Milan *APS* Model 1b. *CAAPS* to blood HbA$_{1c}$ measurements seemed to reflect well performance needs for both diagnostic testing and

monitoring (Table 5). Duplicate or triplicate measurements are needed to monitor disease states. Interestingly, desirable *APS* estimated from CV_I of diabetics for HbA$_{1c}$ (Model 2) gave closely related results despite different background of estimates (Table 5). The *APS* estimated as *CAAPS* (Model 1b) remains our primary choice because of its clinically defined background.

The obvious link to clinical need can be used to guide *APS* for measurements of urine Alb and plasma AmylP (Table 5) because of their key roles in classification of patients with chronic renal disease or acute abdominal pain, respectively.

CAAPS for plasma LDL-C emphasizes the limitations of judging treatment effects with singleton comparison only in case of a wide CV_I. Our calculation provided an estimate of 4.9% from four repeated comparisons until a change of 20% was noted with $1\text{-}\beta = 85\%$ (Table 4). The CV_I based *APS* is 4.0% (Table 5), but nevertheless detection of the prognostic targets expressed in the European guidelines requires four repeated measurements [27]. Thus, both *CAAPS* (Model 1b) and *APS* (Model 2) challenge the feasibility of the current prognostic categories used in dyslipidemia treatment. This highlights the need of clinical guideline developers to involve laboratory professionals in guideline development so that treatment targets are set with clear understanding of the impact of both CV_I and state-of-the-art analytical performance.

Both the between-subject and intra-individual biological variations of plasma Na are narrow in health, resulting in a desirable *APS* = 0.25% for plasma sodium (Table 5). Clinically, the accuracy of plasma Na is critical at concentrations distant from strictly homeostatic health-related reference interval, most importantly in the hyponatremia range. Our modelling provided a *CAAPS* = 0.6% ($1\text{-}\beta = 85\%$), satisfying the need of critically ill patients. We approached incidentally the *TE* estimate = 0.7% from biological variation. All of these were notably tighter than a traditional acceptance limit in EQA schemes [43]. The wide EQA limits emphasize the need to improve technical quality of electrolyte measurements, despite it being a difficult task [46].

Plasma Crea was assigned to Milan Model 2 in the consensus proposal [8], requiring then an *APS* of 2.25% [30]. In clinical practice, renal insufficiency is widely screened and classified using computerized eGFR equations calculated from plasma Crea (Table 1). The frequency of plasma Crea measurements is also explained by need of GFR estimates to avoid overdosing of drugs in renal insufficiency. These examples show that differentiation between healthy and impaired renal function is the key use of plasma Crea measurements, and establish the clinical link required for Model 1b APS. The *CAAPS* of plasma Crea measurements was 8.0% ($1\text{-}\beta = 85\%$), or 5.0% ($1\text{-}\beta = 98\%$), the former being close to 7.4% obtained from biological variation for *TE* (Table 5).

4.3. Statistical flexibility of obtained CAAPS estimates

The *CAAPS* model caters for different z statistics of the equation (2), representing different levels of statistical power $1\text{-}\beta$. It might be used to define optimum, desirable, or minimum *CAAPS* for each measurand, like earlier conventions for $CV_A < 0.25\ CV_I$ (optimum), $<0.5\ CV_I$ (desirable), or $< 0.75\ CV_I$ (minimum) [38,47,48]. A major benefit of the *CAAPS* model is its flexibility if quantitative data are available for classification of disease states at various concentration levels. The *CAAPS* is only as robust as the used variance components, available clinical guidelines, and applied decision limits they are anchored to. Separate variances for chosen decision limits at different concentration levels of the analyte can be modeled if needed [49]. Disease-associated CV_I are larger than the CV_I in health, at least in patients with diabetes or chronic renal failure [50], changing the equation of combined variance between two measurements. Replicate measurements are advisable when using diagnostic cut-offs, as they improve imprecision and consequently accuracy of clinical diagnosis [51]. Furthermore, post-analytical uncertainty was not considered in the proposed model of the *CD*, because it is usually related to discrete non-conformity events rather than increased variance.

4.4. Limitations of the CAAPS approach

4.4.1. Clinical guidelines as sources of APS

Clinically justified analytical performance is essential for clinical laboratory service. However, the complex use of laboratory tests is difficult to translate into needed analytical performance [7]. An *APS* for a single test is not easily isolated from clinical practice with combined tests, other investigations, and other factors of health-care environment. Direct Milan Model 1a studies, i.e., impact of analytical performance of measurements on clinical outcomes, are lacking even with respect to blood HbA$_{1c}$ measurements in diabetes outcomes [9]. The surrogate indirect outcome studies of Milan Model 1b are pragmatic, closely related to clinical classifications and decisions, but they carry inherent limitations, e.g., a tendency to base clinical requirements on state-of-the-art analytical performance [6]. Another limitation is that expert consensus on adequate classification limits may provide a relevant *CD,* and a consequent *APS* for laboratory use, but the impact of use of such *CD* clinically with an analytical performance corresponding to the given *APS* in the laboratory should be verified in a clinical study to show an actual increase in patients' health outcome.

A primary prerequisite for the indirect Milan Model 1b is a well-defined link between clinical decisions and the used test [7]. We derived clinically significant differences, *CD*, from diagnostic categories or prognostic targets from international guidelines to avoid individual opinions (Table 1). In patient monitoring, a difference between two results obtained with optimal analytical performance tends to be perceived always "significant" especially if indicating worsening of a clinical situation, despite the actual effect on patient outcome, as shown with interpretation of blood HbA$_{1c}$ results [21].

Estimates of *CD* for monitoring blood HbA$_{1c}$ were available from long-term clinical outcomes of diabetic patients and international studies (Table 1). With other measurands, we minimized state-of-the-art reasoning by using *CD* from classifications of pathophysiological states (moderate nephropathy, acute pancreatitis), or treatment targets (high-risk dyslipidemia), independently of assay performance. If an accurate *CD* was missing, a consensus of the authors was used to define the quantitative limits of each *CD*. In this way, the moderate albuminuria was described by the logarithmic midpoint of its range (100 mg/L urine Alb). For plasma AmylP, a sensitized limit from 4xURL to 2xURL was used to improve detection of delayed or severe pancreatitis, based on sensitivity Sn and specificity Sp of elevated plasma AmylP in the diagnostics of pancreatitis. These two midpoint examples show tailoring of the *CD* criteria to specific clinical purposes [7].

In profound hyponatremia, the clinical need of accuracy is associated with the risk of cerebral edema resulting from too rapid correction of plasma Na (Table 1). Hyponatremia treatment protocol from the European hyponatremia guideline quotes the same critical difference (125–130 mmol/L) as used earlier by Klee in reporting "medical utility *CV*" for plasma Na [52]. Clinical knowledge may be universal if relevant clinical situations for *CD* estimates are selected, although confirmation from local clinicians is always recommended.

4.4.2. Limits of modeling acceptable CV_A from estimates of CV_D and its components

To model *APS* from clinically significant differences, *CD*, the concept of reference change value *RCV* [28,34] was applied from existing examples for measurements of HbA$_{1c}$ [19,21,9]. The guideline-derived *CD* (=RCV) was used to calculate maximum acceptable diagnostic variation CV_D (equation (2)) that was further divided into its components CV_I, CV_{PRE} and CV_A (equation (3)), to reach *CAAPS* (Fig. 1). The modelling assumes Gaussian distributions of *CV* data with identical variances in repeated measurements. We then simplified the published non-Gaussian distributions of health related CV_I using median values as means [30,41]. To improve sensitivity in detecting a true change from 50% (corresponding $1\text{-}\beta$ at $z = 1.96$, $\alpha = 2.5\%$), we increased the z score to 3 or 4 [29] (Fig. 2).

A practically important problem is a large CV_I of some measurands that exhausts the budget available for CV_A from a relatively narrow CV_D in subtractions, as shown for plasma LDL-C (Table 2). We offer repeated measurements as an option to reach *CAAPS* with these measurands (Table 4). An average CV_I of ambulatory patients may also be reduced in standardized environments using short collection intervals from in-patients (emergency room, intensive care), to detect a tendency of repeated results better, such as that in plasma troponin I concentrations, when suspecting of cardiac events. Occasionally, another measurand with a smaller CV_I may solve the need for improved diagnostics.

5. Conclusions

Clinically significant differences *CD* were searched from international guidelines and modelled into two measurements to enable calculating diagnostic variation CV_D with traditional *RCV* statistics and deriving *CAAPS* for six clinical chemistry measurands representing different areas of clinical diagnostics. The *CAAPS* provides a new tool to anchor laboratory performance to clinical needs using well-defined settings. The calculations can be further adjusted for different levels of required statistical power and repeated measures making them applicable to a wide range of clinical scenarios and analytical performance. Thus, the *CD*-derived *CAAPS* support granular discussions on test performance between administration, clinicians, and laboratories instead of providing *APS* as external facts that professionals outside the laboratories cannot challenge.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Calculations used for Table 4 were shared as a Supplementary Excel table for the readers

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.cca.2023.117233.

References

[1] R.N. Barnett, Medical significance of laboratory results, Am. J. Clin. Pathol. 50 (1968) 671–677.

[2] E. Cotlove, E.K. Harris, G.Z. Williams, Biological and analytic components of variation in long-term studies of serum constituents in normal subjects. III Physiological and medical implications, Clin. Chem. 16 (1970) 1028–1032.

[3] D.B. Tonks, M. Stoeppler, D.E. Douglas, J. Wolska, Evaluation of analytical techniques for measurements of cadmium in body fluids, Ann. Clin. Lab. Sci. 15 (1985) 342–342.

[4] C.G. Fraser, P.H. Petersen, Analytical performance characteristics should be judged against objective quality specifications, Clin. Chem. 45 (1999) 321–323. PMID: 10053031.

[5] D. Kenny, C.G. Fraser, P. Hyltoft Petersen, A. Kallner, Consensus agreement, Scand. J. Clin. Lab. Invest. 59 (1999) 585–585, doi: 10.1080/00365519950185409.

[6] S. Sandberg, C.G. Fraser, A.R. Horvath, R. Jansen, G. Jones, W. Oosterhuis, 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. 53 (2015) 833–835, https://doi.org/10.1515/cclm-2015-0067.

[7] A.R. Horvath, P.M. Bossuyt, S. Sandberg, A.S. John, P.J. Monaghan, W.D. Verhagen-Kamerbeek, et al. for the Test Evaluation Working Group of the European Federation of Clinical Chemistry and Laboratory Medicine, Setting analytical performance specifications based on outcome studies - is it possible? Clin. Chem. Lab. Med. 53 (2015) 841–848, https://doi.org/10.1515/cclm-2015-0214.

[8] F. Ceriotti, P. Fernandez-Calle, G.G. Klee, G. Nordin, S. Sandberg, T. Streichert, et al., Criteria for assigning laboratory measurands to models for analytical performance specifications defined in the 1st EFLM Strategic Conference, Clin. Chem. Lab. Med. 55 (2017) 189–194, https://doi.org/10.1515/cclm-2016-0091.

[9] T.P. Loh, A.F. Smith, K.J.L. Bell, S.J. Lord, F. Ceriotti, G. Jones, et al., Setting analytical performance specifications using HbA1c as a model measurand, Clin. Chim. Acta 523 (2021) 407–414, https://doi.org/10.1016/j.cca.2021.10.016.

[10] Joint Committee for Guides in Metrology, International vocabulary of metrology — Basic and general concepts and associated terms, third ed., VIM 3, 2012. Available from: <https://www.bipm.org/utils/common/documents/jcgm/JCGM_200_2012.pdf>. (accessed 20 July, 2022).

[11] International Organization for Standardization, ISO/TS 20914:2019, Medical laboratories — practical guidance for the estimation of measurement uncertainty, International Organization for Standardization, Geneva, 2019.

[12] International Organization for Standardization, ISO 17511:2020, In vitro diagnostic medical devices — requirements for establishing metrological traceability of values assigned to calibrators, trueness control materials and human samples, International Organization for Standardization, Geneva, 2020.

[13] E. Theodorsson, Uncertainty in measurement and total error: tools for coping with diagnostic uncertainty, Clin. Lab. Med. 37 (2017) 15–34, https://doi.org/10.1016/j.cll.2016.09.002.

[14] T. Kouri, M. Siloaho, S. Pohjavaara, P. Koskinen, O. Malminiemi, P. Pohja-Nylander, R. Puukka, Pre-analytical factors and measurement uncertainty, Scand J Clin Lab Invest 65 (2005) 463–475, https://doi.org/10.1080/00365510500208332.

[15] B. Magnusson, H. Ossowicki, O. Rienitz, E. Theodorsson, Routine internal- and external-quality control data in clinical laboratories for estimating measurement and diagnostic uncertainty using GUM principles, Scand. J. Clin. Lab. Invest. 72 (2012) 212–220, https://doi.org/10.3109/00365513.2011.649015.

[16] REGULATION (EU) 2017/746 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 5 April 2017 on in vitro diagnostic medical devices and repealing Directive 98/79/EC and Commission Decision 2010/227/EU. Available from: <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32017R0746&qid=1648147617108>.

[17] International Organization for Standardization, ISO 15189:2022, Medical laboratories - requirements for quality and competence, third ed., International Organization for Standardization, Geneva, 2022.

[18] World Health Organization, Classification of diabetes mellitus, World Health Organization, Geneva, 2019.

[19] R.R. Little, C.L. Rohlfing, D.B. Sacks, for the National Glycohemoglobin Standardization Program (NGSP) Steering Committee, Status of hemoglobin A(1c) measurement and goals for improvement: from chaos to order for improving diabetes care, Clin. Chem. 57 (2011) 205–214, https://doi.org/10.1373/clinchem.2010.148841.

[20] C. Weykamp, HbA1c: a review of analytical and clinical aspects, Ann. Lab. Med. 33 (2013) 393–400, https://doi.org/10.3343/alm.2013.33.6.393.

[21] S. Skeie, C. Perich, C. Ricos, A. Araczki, A.R. Horvath, W.P. Oosterhuis, et al., Postanalytical external quality assessment of blood glucose and hemoglobin A1c: an international survey, Clin. Chem. 51 (2005) 1145–1153, https://doi.org/10.1373/clinchem.2005.048488.

[22] R.C. Turner, R.R. Holman, C.A. Cull, I.M. Stratton, D.R. Matthews, V. Frighi, et al., UK Prospective Diabetes Study (UKPDS) Group, Intensive blood-glucose control with sulphonylureas or insulin compared with conventional treatment and risk of complications in patients with type 2 diabetes (UKPDS 33), Lancet 352 (1998) 837–853, https://doi.org/10.1016/S0140-6736(98)07019-6.

[23] Kidney Disease: Improving Global Outcomes (KDIGO) CKD Work Group. KDIGO 2012 Clinical Practice Guideline for the Evaluation and Management of Chronic Kidney Disease. Kidney Int, Suppl. 3 (2013) 1–150. Summary of Recommendation Statements. doi: 10.1038/kisup.2012.77.

[24] S.S. Waikar, C.M. Rebholz, Z.H. Zheng, S. Hurwitz, C.Y. Hsu, H.I. Feldman, et al., Chronic kidney disease biomarkers consortium, biological variability of estimated GFR and albuminuria in CKD, Am. J. Kidney Dis. 72 (2018) 538–546, https://doi.org/10.1053/j.ajkd.2018.04.023.

[25] G. Spasovski, R. Vanholder, B. Allolio, D. Annane, S. Ball, D. Bichet, et al., on behalf of the Hyponatraemia Guideline Development Group, Clinical practice guideline on diagnosis and treatment of hyponatraemia, Eur. J. Endocrinol. 170 (2014) G1–G47, https://doi.org/10.1530/EJE-13-1020.

[26] S. Tenner, J. Baillie, J. DeWitt, S.S. Vege, American college of gastroenterology guideline: management of acute pancreatitis, Am. J. Gastroenterol. 108 (2013) 1400–1415, https://doi.org/10.1038/ajg.2013.218.

[27] F. Mach, C. Baigent, A.L. Catapano, K.C. Koskinas, M. Casula, L. Badimon, et al., the Task Force for the management of dyslipidaemias of the European Society of Cardiology (ESC) and European Atherosclerosis Society (EAS), 2019 ESC/EAS Guidelines for the management of dyslipidaemias: lipid modification to reduce cardiovascular risk, Eur. Heart J. 41 (2020) 111–188, doi: 10.1093/eurheartj/ehz455.

[28] C.G. Fraser, Reference change values, Clin. Chem. Lab. Med. 50 (2012) 807–812, https://doi.org/10.1515/cclm.2011.733.

[29] N. Iglesias Canadell, P. Hyltoft Petersen, E. Jensen, C. Ricos, P.E. Jorgensen, Reference change values and power functions, Clin. Chem. Lab. Med. 42 (2004) 415–422, https://doi.org/10.1515/CCLM.2004.073.

[30] A.K. Aarsand, P. Fernandez-Calle, C. Webster, A. Coskun, E. Gonzales-Lao, J. Diaz-Garzon, et al., on behalf of the European Federation of Clinical Chemistry and Laboratory Medicine Task and Finish Group for the Biological Variation Database, The EFLM Biological Variation Database, 2022. Available from: <https://biologicalvariation.eu/> (accessed 3 November 2022).

[31] M.S. Sylte, T. Wentzel-Larsen, B.J. Bolann, Estimation of the minimal preanalytical uncertainty for 15 clinical chemistry serum analytes, Clin. Chem. 56 (2010) 1329–1335, https://doi.org/10.1373/clinchem.2010.146050.

[32] X. Fuentes-Arderiu, G. Acebes-Frieyro, L. Gavaso-Navarro, M.J. Castiñeiras-Lacambra, Pre-metrological (pre-analytical) variation of some biochemical quantities, Clin. Chem. Lab. Med. 37 (1999) 987–989, https://doi.org/10.1515/CCLM.1999.146.

[33] R. Rigo-Bonnin, D. Munoz-Provencio, F. Canalias, Reference change values based on uncertainty models, Clin. Biochem. 80 (2020) 31–41, https://doi.org/10.1016/j.clinbiochem.2020.03.016.

[34] E.K. Harris, T. Yasaka, On the calculation of a "reference change" for comparing two consecutive measurements, Clin. Chem. 29 (1983) 25–30.

[35] N. Iglesias, P. Hyltoft Petersen, C. Ricos, Power function of the reference change value in relation to cut-off points, reference intervals and index of individuality, Clin. Chem. Lab. Med. 43 (2005) 441–448, https://doi.org/10.1515/CCLM.2005.078.

[36] J.O. Westgard, T. Groth, T. Aronsson, H. Falk, C.H. de Verdier, Performance characteristics of rules for internal quality control: probabilities for false rejection and error detection, Clin. Chem. 23 (1977) 1857–1867.

[37] C.G. Fraser, Test result variation and the quality of evidence-based clinical guidelines, Clin. Chim. Acta 346 (2004) 19–24, https://doi.org/10.1016/j.cccn.2003.12.032.

[38] F. Braga, M. Panteghini, The utility of measurement uncertainty in medical laboratories, Clin. Chem. Lab. Med. 58 (2020) 1407–1413, https://doi.org/10.1515/cclm-2019-1336.

[39] A.A. Nielsen, P.H. Petersen, A. Green, C. Christensen, H. Christensen, I. Brandslund, Changing from glucose to HbA1c for diabetes diagnosis: predictive values of one test and importance of analytical bias and imprecision, Clin. Chem. Lab. Med. 52 (2014) 1069–1077, https://doi.org/10.1515/cclm-2013-0337.

[40] F. Braga, M. Panteghini, Performance specifications for measurement uncertainty of common biochemical measurands according to Milan models, Clin. Chem. Lab. Med. 59 (2021) 1362–1368, https://doi.org/10.1515/cclm-2021-0170.

[41] S. Carlsen, P. Hyltoft Petersen, S. Skeie, Ø. Skadberg, S. Sandberg, Within-subject biological variation of glucose and HbA1c in healthy persons and in type 1 diabetes patients, Clin. Chem. Lab. Med. 49 (2011) 1501–1507.

[42] L.M. Bachmann, G. Nilsson, D.E. Bruns, M.J. McQueen, J.C. Lieske, J.J. Zakowski, W.G. Miller, State of the art for measurement of urine albumin: comparison of routine measurement procedures to isotope dilution tandem mass spectrometry, Clin. Chem. 60 (2014) 471–480, https://doi.org/10.1373/clinchem.2013.210302.

[43] S. Westgard, Consolidated Comparison of Chemistry (and Toxicology) Performance Specifications, 2022. Available from: <https://www.westgard.com/consolidated-goals-chemistry.htm> (accessed 22 July 2022).

[44] B.C. Birkelo, N. Pannu, E.D. Siew, Overview of diagnostic criteria and epidemiology of acute kidney injury and acute kidney disease in the critically ill patient, Clin. J. Am. Soc. Nephrol. 17 (2022) 717–735, https://doi.org/10.2215/CJN.14181021.

[45] G.R.D. Jones, Further issues with using reference change values, Clin. Chim. Acta 528 (2022) 13–14, https://doi.org/10.1016/j.cca.2022.01.008.

[46] S. Pasqualetti, M. Chibireva, F. Borrillo, F. Braga, M. Panteghini, Improving measurement uncertainty of plasma electrolytes: a complex but not impossible task, Clin. Chem. Lab. Med. 59 (2021) e129–e132, https://doi.org/10.1515/cclm-2020-1399.

[47] C.G. Fraser, P. Hyltoft Petersen, J.C. Libeer, C. Ricos, Proposals for setting generally applicable quality goals solely based on biology, Ann. Clin. Biochem. 34 (1997) 8–12, https://doi.org/10.1177/000456329703400103.

[48] M. Panteghini, F. Ceriotti, G. Jones, W. Oosterhuis, M. Plebani, S. Sandberg, on behalf of the Task Force on Performance Specifications in Laboratory Medicine of the European Federation of Clinical Chemistry and Laboratory Medicine (EFLM), Strategies to define performance specifications in laboratory medicine: 3 years on from the Milan Strategic Conference, Clin. Chem. Lab. Med. 55 (2017) 1849–1856, https://doi.org/10.1515/cclm-2017-0772.

[49] G.R.D. Jones, Critical difference calculations revised: inclusion of variation in standard deviation with analyte concentration, Ann. Clin. Biochem. 46 (2009) 517–519, https://doi.org/10.1258/acb.2009.009083.

[50] C. Ricos, N. Iglesias, J.V. Garcia-Lario, M. Simon, F. Cava, A. Hernandez, et al., Within-subject biological variation in disease: collated data and clinical consequences, Ann. Clin. Biochem. 44 (2007) 343–352, https://doi.org/10.1258/000456307780945633.

[51] P.H. Petersen, G.G. Klee, Influence of analytical bias and imprecision on the number of false positive results using Guideline-Driven Medical Decision Limits Clin. Chim. Acta 430 (2014) 1–8, https://doi.org/10.1016/j.cca.2013.12.014.

[52] G.G. Klee, Establishment of outcome-related analytic performance goals, Clin. Chem. 56 (2010) 714–722, https://doi.org/10.1373/clinchem.2009.133660.