NICOLAUS COPERNICUS UNIVERSITY IN TORUŃ
Faculty of Pharmacy
Collegium Medicum in Bydgoszcz

INTERNATIONAL SCIENTIFIC MEETING "EVALUATING AND MONITORING ANALYTICAL QUALITY IN THE TRACEABILITY ERA"

29 NOVEMBER 2024

Analytical performance specifications: moving from theoretical principles to practice

Mauro Panteghini

EVERY DAY IS
PATIENT SAFETY
DAY

Variability between methods

↓

Incorrect patient results

↓

Mis-diagnosis / mis-management

↓

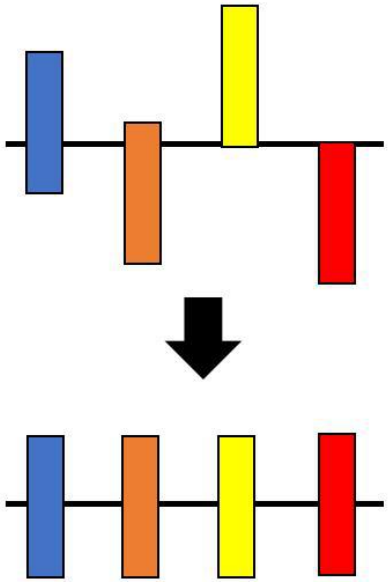
Poor clinical outcomes

↓

Impact on patient safety

The world population of

8 billion people

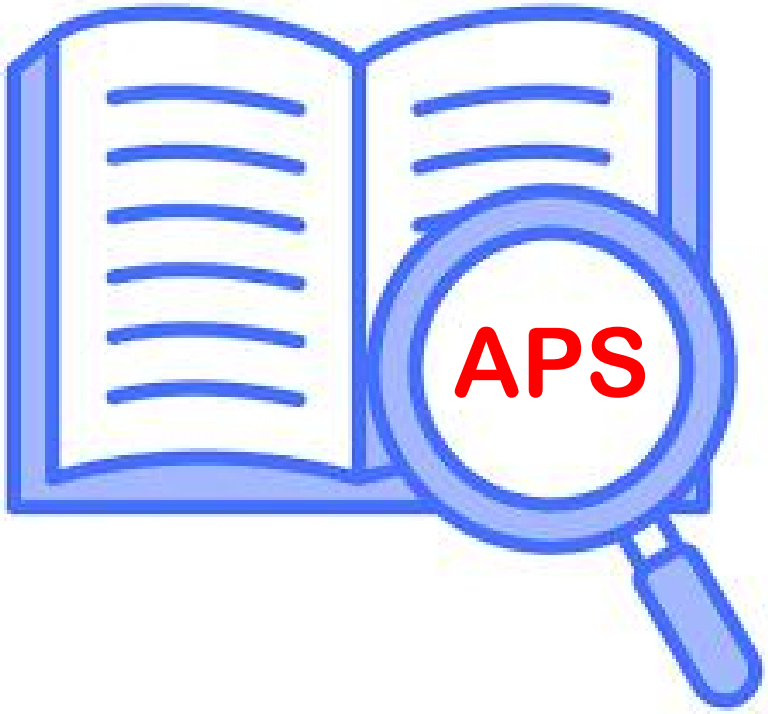


is entitled to believe that all methods will give the same result on their specimen

Obtaining harmonization of laboratory results is an absolute priority for public health

Real Sense

The goal of standardization is “accurate measurement”, which is measurement associated with an *acceptable* measurement uncertainty from a recognized standard

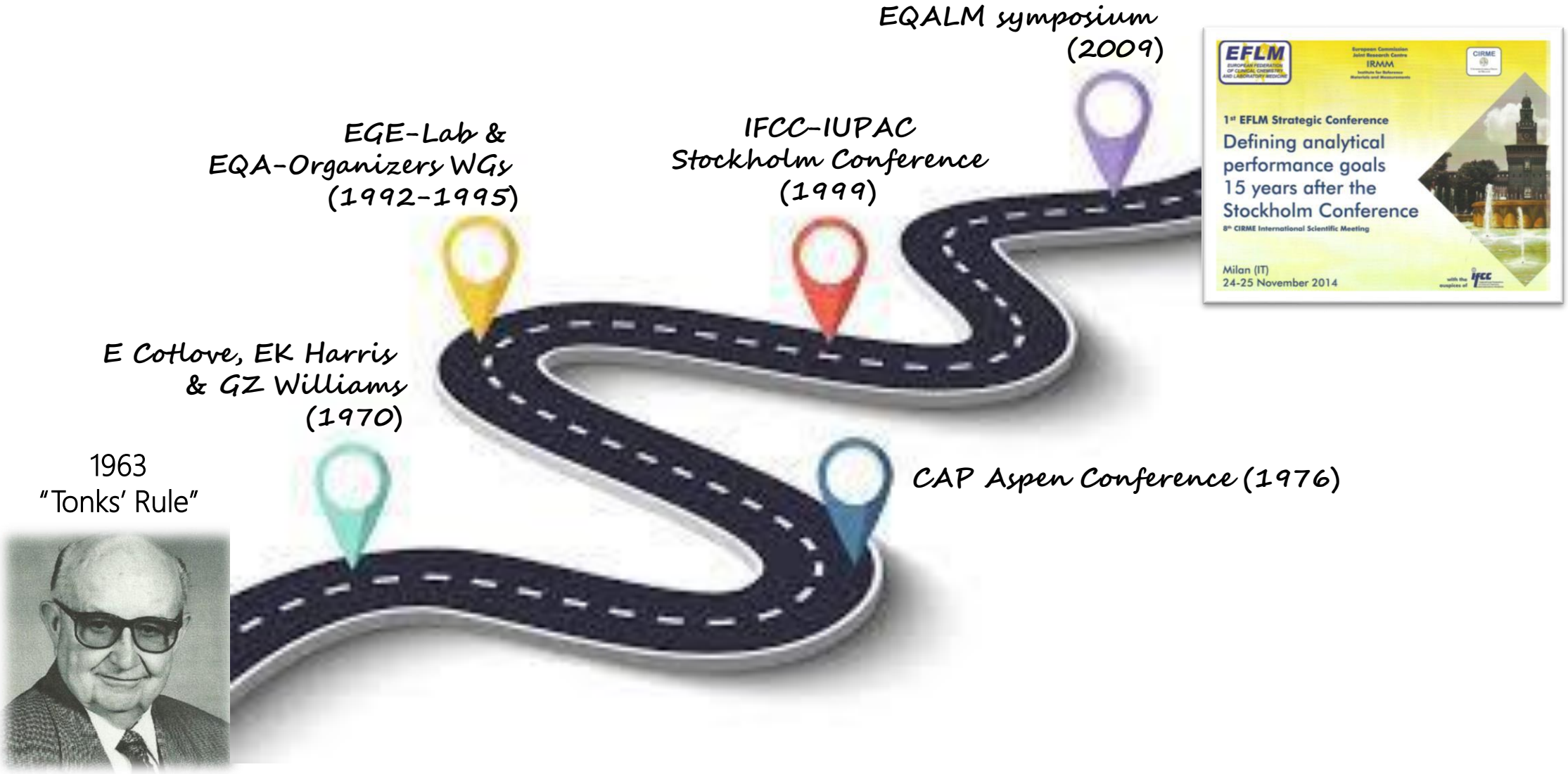


Analytical performance specifications

Set of criteria that specify (in numerical terms) the quality required for values assigned to a clinical sample to satisfy clinical needs

If APS are not objectively defined and fulfilled, the variation in laboratory result may overwhelm the clinical information supplied, potentially causing a negative effect on patient's outcome

Setting analytical performance specifications: "a long and winding road"



The 2014 Milan Strategic Conference heritage

Three models to set APS based on the effects on patients' outcome, biological variation of the measurand, and state of the art of the measurements, respectively
The models rely on different principles and do not constitute a hierarchy. Accordingly, some models are better suited for certain measurands than for others, and the attention for model selection should primarily direct toward the measurand and its biological and clinical characteristics
APS can be different for different applications of the same test, but if a test is used for multiple purposes the strictest APS should take precedence
APS should derive from high quality studies and updated data
Grading different quality levels of APS is helpful to define priorities in ameliorating actions

[Adapted from Panteghini M. Clin Chem Lab Med 2024;62:1455]

DE GRUYTER

Clin Chem Lab Med 2017; 55(2): 189–194

Opinion Paper

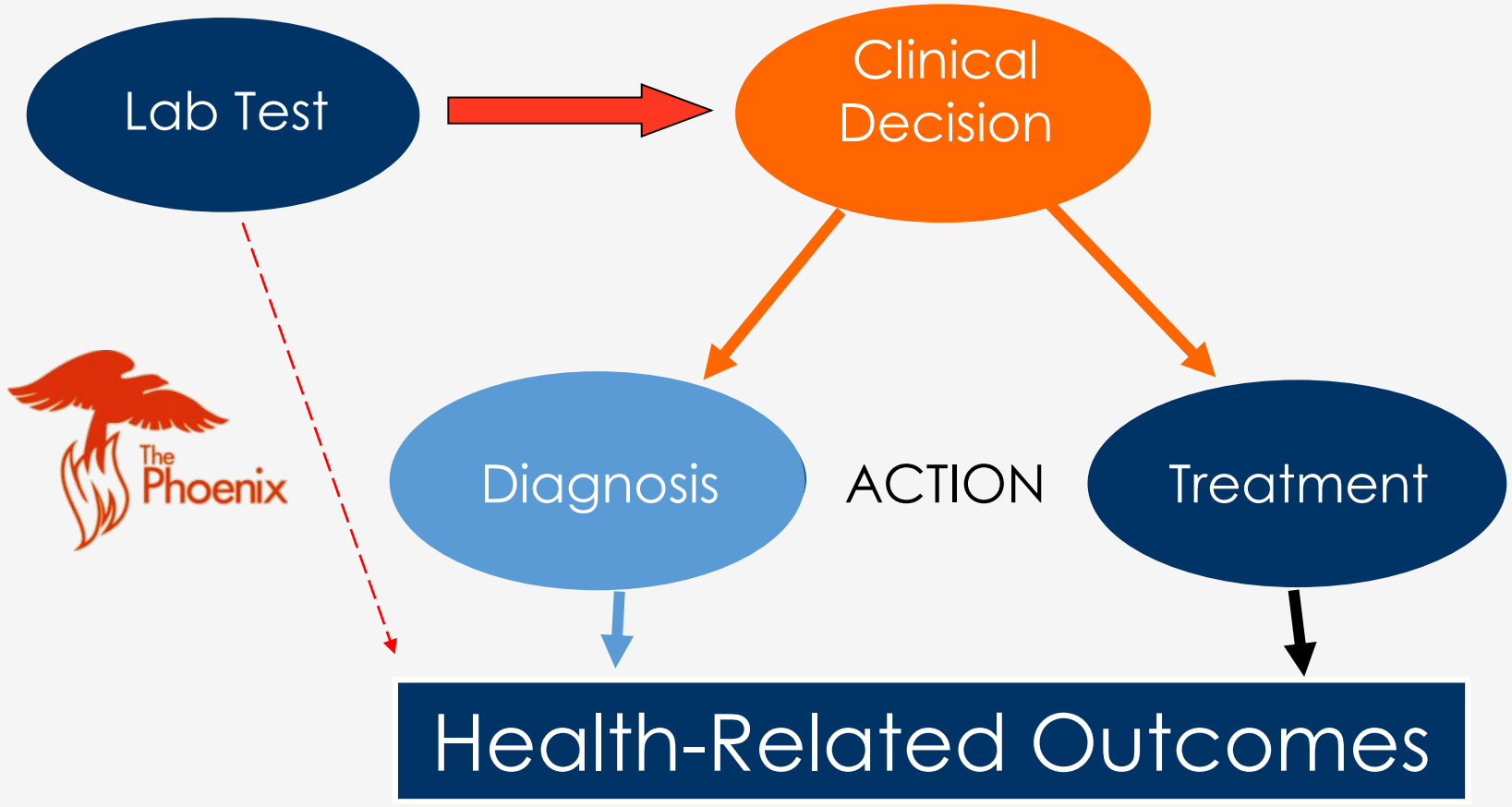
Ferruccio Ceriotti*, Pilar Fernandez-Calle, George G. Klee, Gunnar Nordin, Sverre Sandberg, Thomas Streichert, Joan-Lluis Vives-Corrons and Mauro Panteghini, on behalf of the EFLM Task and Finish Group on Allocation of laboratory tests to different models for performance specifications (TFG-DM)

Criteria for assigning laboratory measurands to models for analytical performance specifications defined in the 1st EFLM Strategic Conference

Model 1: Based on the effect of analytical performance on clinical outcome

- Applied to the measurands with a central and well-defined role in the diagnosis of a specific disease or a given clinical situation, with test results being interpreted through established common decision thresholds.

Challenge: Directly Connecting Laboratory Testing to Outcome



George Box, 1919-2013

"The best models are not necessarily the most useful models"

"Concerning the outcome-based model, everybody at the conference agrees that using direct outcome studies for defining APS, although it represents the ideal approach, is however impossible to be translated into practical use."

[Sandberg S, Zima T, Panteghini M. Clin Chem Lab Med 2024;62:1451]

Defining APS using *indirect* outcome data

- ✓ Consider the impact of analytical performance of the test on clinical (mis)classifications or decisions and thereby on probability of outcomes, using simulation or decision analysis.
- ✓ Studies have employed contour plots to present findings from which to derive APS information according to a given rate of clinical misclassification.
- ✓ Computerized approaches using the framework of simulation studies have also become available.

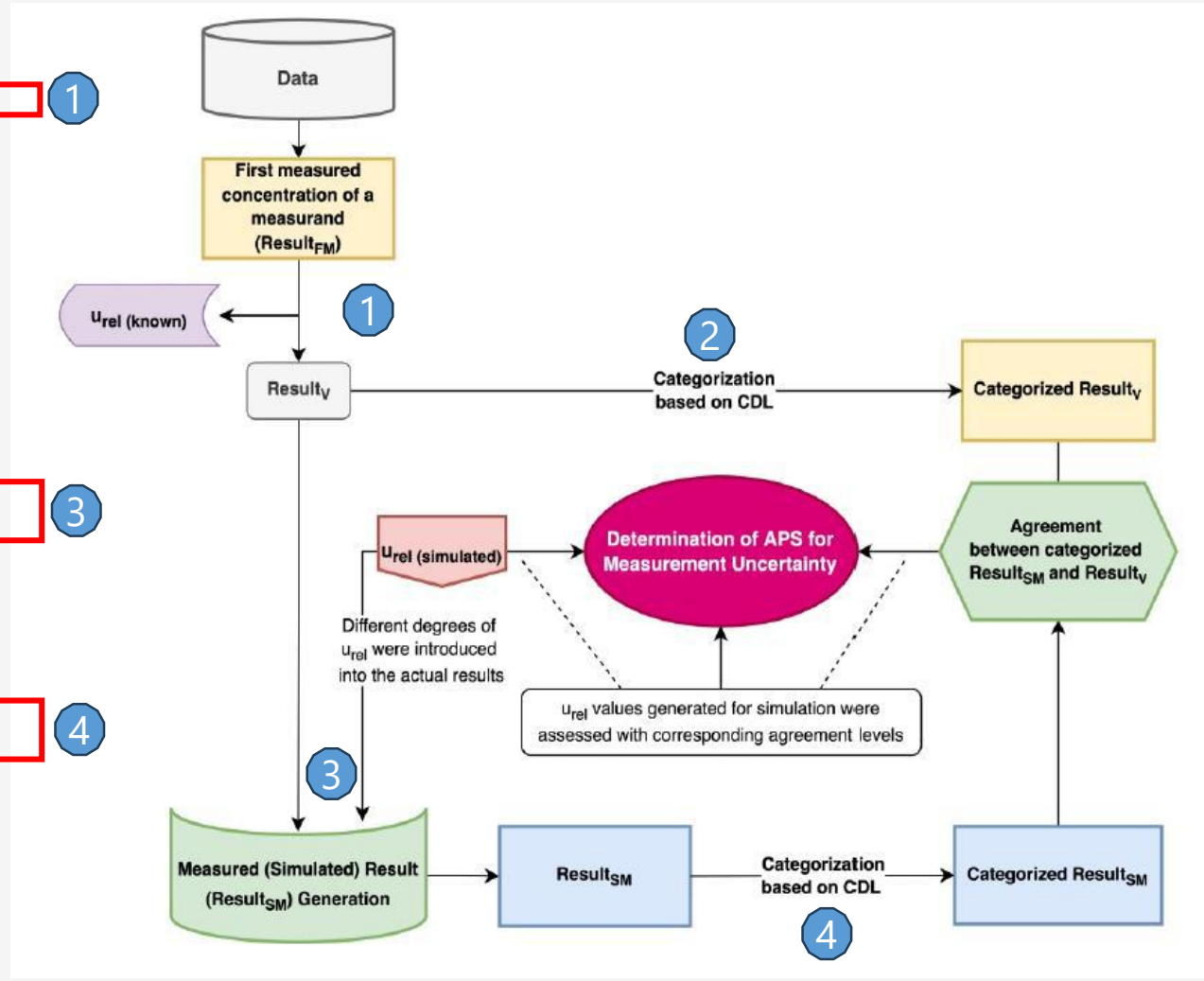
New computerised approaches using the framework of simulation studies

Analytical framework	Key methods identified
Step 1: Calculation of "true" test values	❖ **Empirical data values:** i. Values based on method comparison/EQA studies ii. Empirical values used within nonempirical approaches outlined in Step 2 ❖ **Simulated values:** i. Series of fixed values ii. Parametric distribution (e.g., uniform, gaussian, log-gaussian) iii. Nonparametric distribution (e.g., sampling with replacement)
Step 2: Calculation of measured test values (incorporating measurement uncertainty)	❖ **Empirical data values:** i. Values based on method comparison/EQA studies ❖ **Graphical assessment:** i. Cumulative % frequencies plotted on the probit scale: bias/imprecision applied by shifting/rotating the plotted lines ❖ **Simulation:** i. Iterative simulation according to a specified error model ii. Simulation of uncertain distributions around individual points, or at a population level ❖ **Regression analysis:** i. Uncertainty applied via adjusting "true" values within the regression model (e.g., multiplicative factor)
Step 3: Calculation of the impact of measurement discrepancies on outcomes	❖ **Clinical accuracy:** i. Observe the change in misclassification rates ❖ **Clinical utility (impact on treatment decisions):** i. Test values matched to treatment protocols ii. Error grid analysis ❖ **Costs:** i. Cost estimates applied to misclassification rates ❖ **Clinical utility (health outcomes) and cost-effectiveness:** i. Decision analytic modeling

1

3

4



APS Calculator

A Data-Driven Tool for Setting Outcome-Based Analytical Performance Specifications for Measurement Uncertainty Using Specific Clinical Requirements and Population Data

Serum HDL cholesterol and triglyceride APS
using National Health and Nutrition Examination Survey (NHANES) data series and ATP III decision limits (CDL)

ATP III Serum Triglyceride Classification

	mg/dL
• Physiologic	<150
• Borderline high	150–199
• High	200–499
• Very high	≥500

ATP III Serum HDL Chol Classification

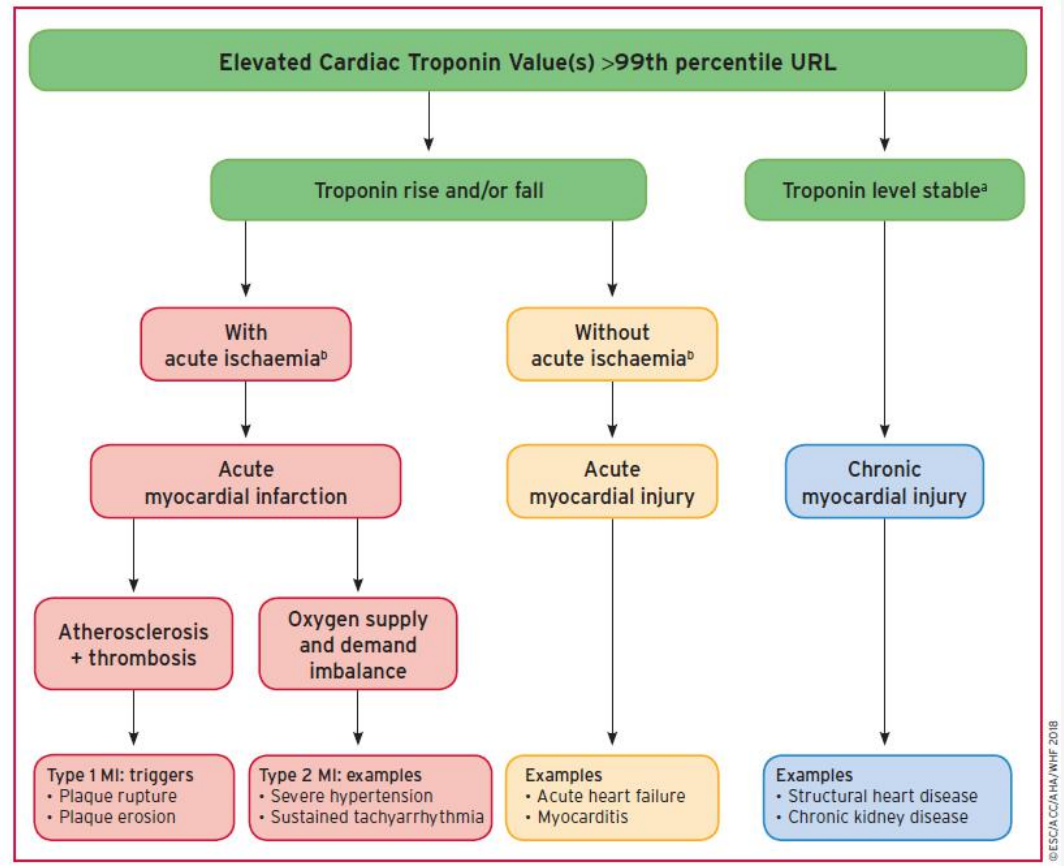
• Low	<40
• Desirable	40–59
• High	≥60

Measurand	APS for standard MU by APS calculator	
	Desirable (defined as 95% agreement)	Minimum (defined as 90% agreement)
S-HDL cholesterol	2.9%	5.6%
S-Triglycerides	6.1%	12.4%

[Çubukçu HC et al. Clin Chem Lab Med 2024;62:597]

Cardiac troponin typically possesses characteristics of measurands that should be allocated to the outcome-based model for deriving APS, i.e. it have a central role in decision-making regarding a specific clinical situation (acute coronary syndrome), with results interpreted through established criteria [an increase of troponin values >99th percentile limit of the reference distribution (upper reference limit)]

EXAMPLE

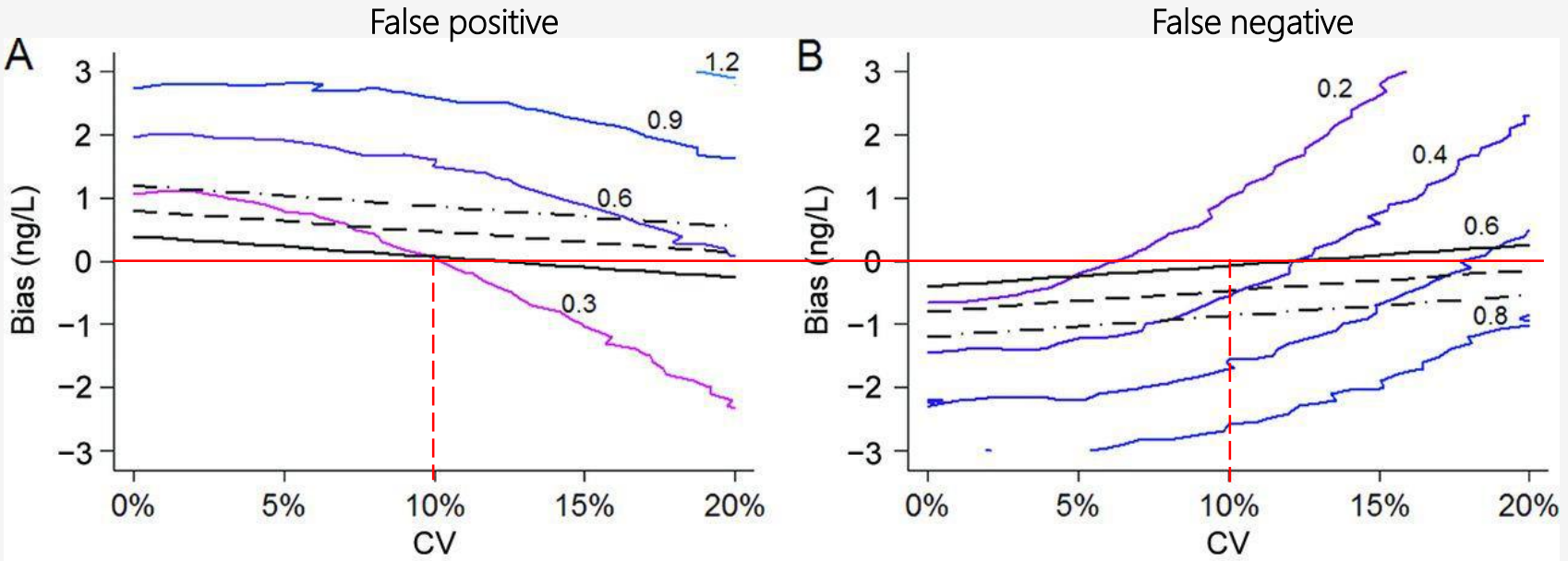


Diagnosis

Fourth Universal Definition of Myocardial Infarction (2018)

Studying the effects of varying analytical performance of cardiac troponin on the acute myocardial infarction diagnosis

Plots of the fraction of hs-TnI misclassification rate as a function of assay performance at the 99th percentile upper reference limit

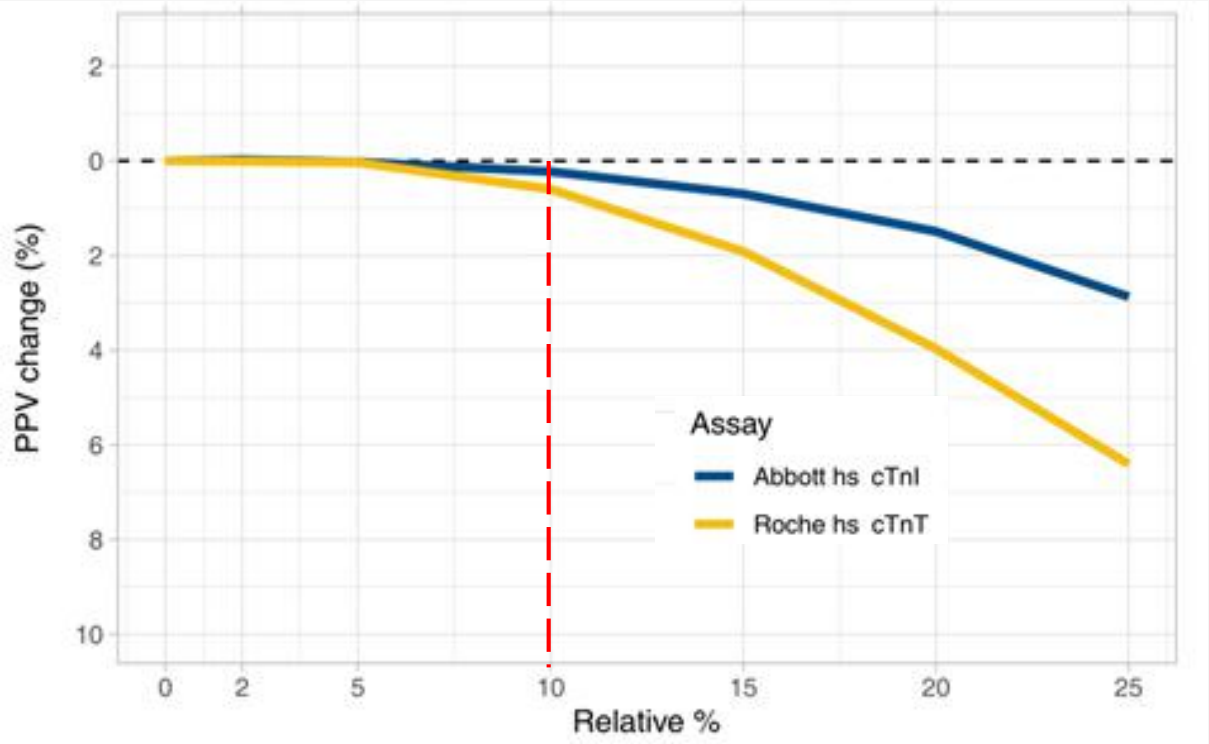


The rate of misclassification was approximately 0.3% at CV 10% and zero bias

[Lyon AW et al. Clin Chem 2017;63:585]

Studying the effects of varying analytical performance of cardiac troponin on the acute myocardial infarction diagnosis

Plot of the hs-TnI and hs-TnT PPV rates as a function of assay performance at the 99th percentile upper reference limit



The rate of misclassification was ≤0.3% at 10% analytical variation

[Pickering JW et al. Clin Chem 2024;70:967]

Facts about APS for troponin measurements

1. [Indirect] outcome studies have focused on the dichotomic clinical (mis)classification of patients with suspected AMI by applying the assay 99th percentile URL
2. These studies indicate that a standard measurement uncertainty <10% @URL may represent a suitable goal as it may maintain the misclassification rate below 0.5%.

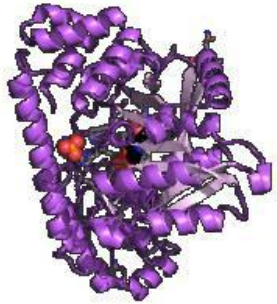
Examples of APS for standard measurement uncertainty using *indirect* outcome approaches

Measurand	u_{result} APS, %	
	Desirable	Minimum
Outcome-based model		
Fasting plasma glucose	2.00	3.00
Blood HbA_{1c}	3.00	3.70
Blood total hemoglobin	5.60	8.50
Serum total cholesterol	3.00	7.00
Serum HDL cholesterol	2.90	5.60
Serum triglycerides	6.10	12.4
Serum cardiac troponin	9.40	13.0
Urine albumin	9.00	17.0
Serum total folate	8.00	12.0
Serum 25-hydroxyvitamin D_3	10.0	15.0
Serum transferrin saturation	10.0	15.0

[Panteghini M. Clin Chem Lab Med 2024;62:1497]

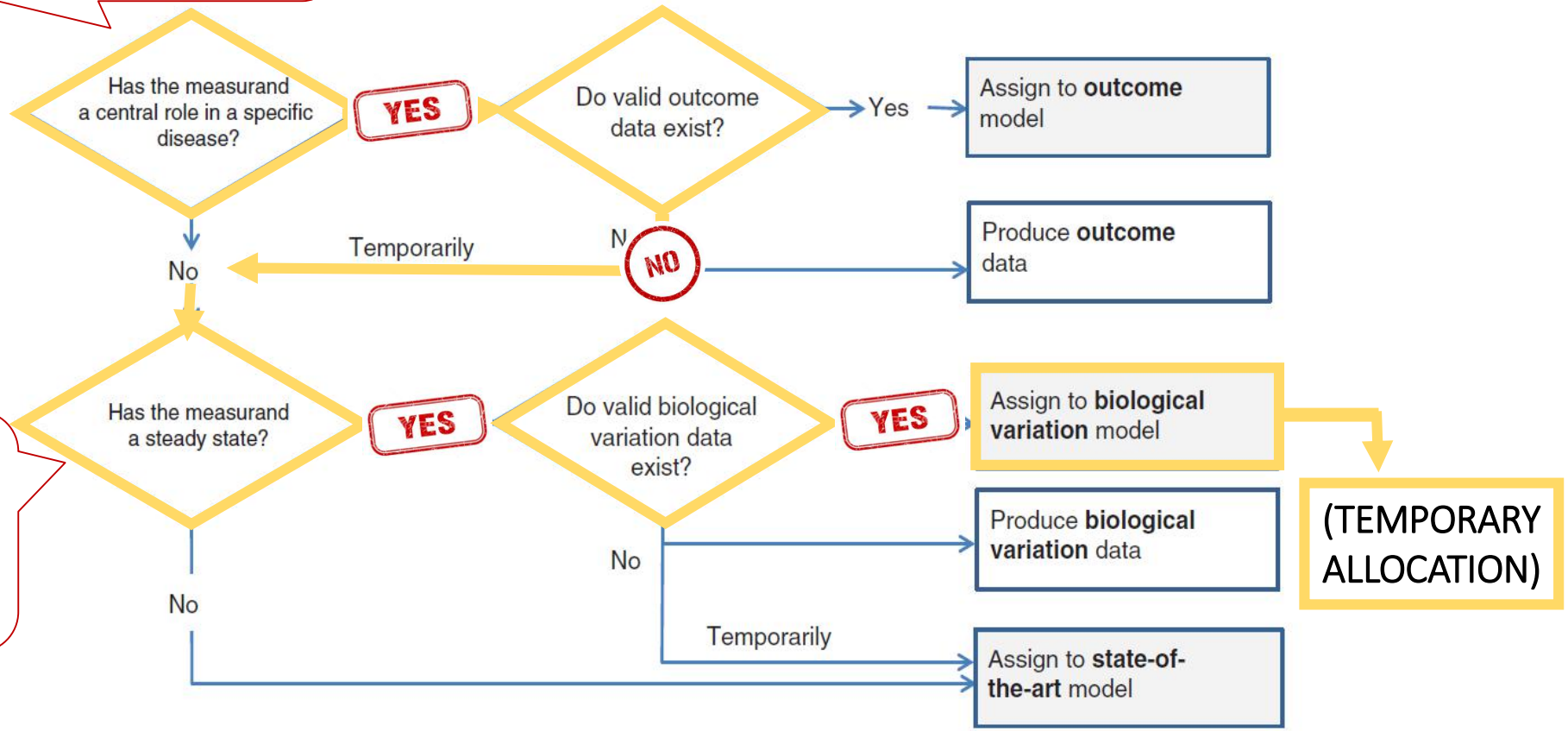
What to do in the absence of information about outcome-based APS for measurands that theoreticaly should be allocated to this model?

→ Temporary allocations to one of the other two Milan models should be considered, according to the measurand characteristics

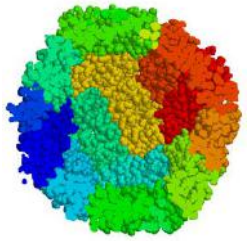


Serum ALT is the first-level test to detect hepatopathies

However, no outcome-based data are available

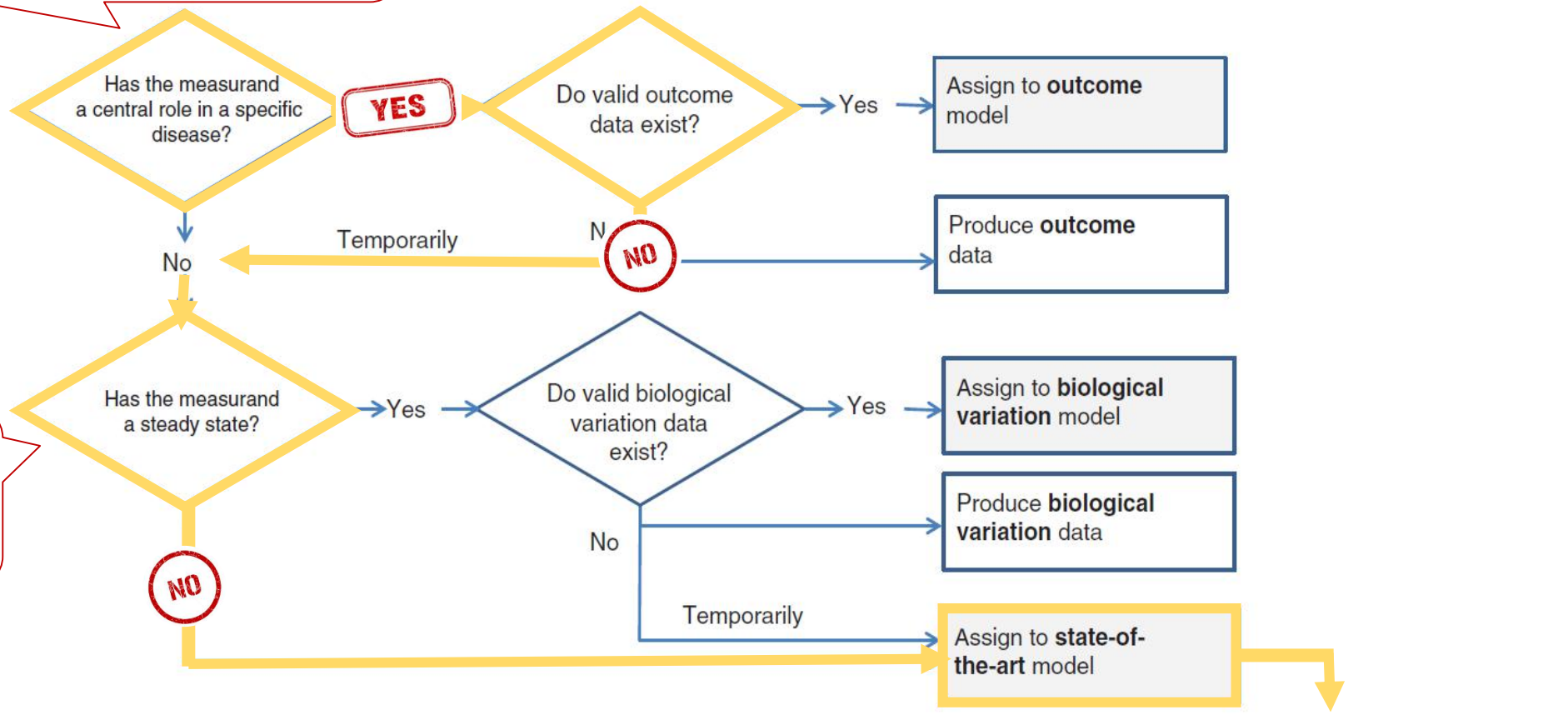


Serum ALT demonstrates rather stable activities in healthy individuals



Serum ferritin is used in the decision-making process of iron-related disorders

However, no outcome-based data are available



Serum ferritin is a sensitive acute phase reactant

(TEMPORARY ALLOCATION)

Model 2: Based on biological variation of the measurand

- Applied to measurands with high homoeostatic control (e.g., plasma ions) or if a measurand has a de facto stable concentration when a subject is in good health (e.g., serum creatinine).
- Should not be used for measurands having insufficient steady state status, such as, for instance, some hormones and urine parameters.
- Therefore, it is not acceptable to use indiscriminately the BV-based model to derive APS for all measurands instead of filtering only measurands that should be allocated to this model.

Concepts behind setting APS from biological variation

$$V_{TOT} = (MU^2 + CV_I^2)^{1/2}$$

analytical variability [measurement uncertainty]

intra-individual biological variability

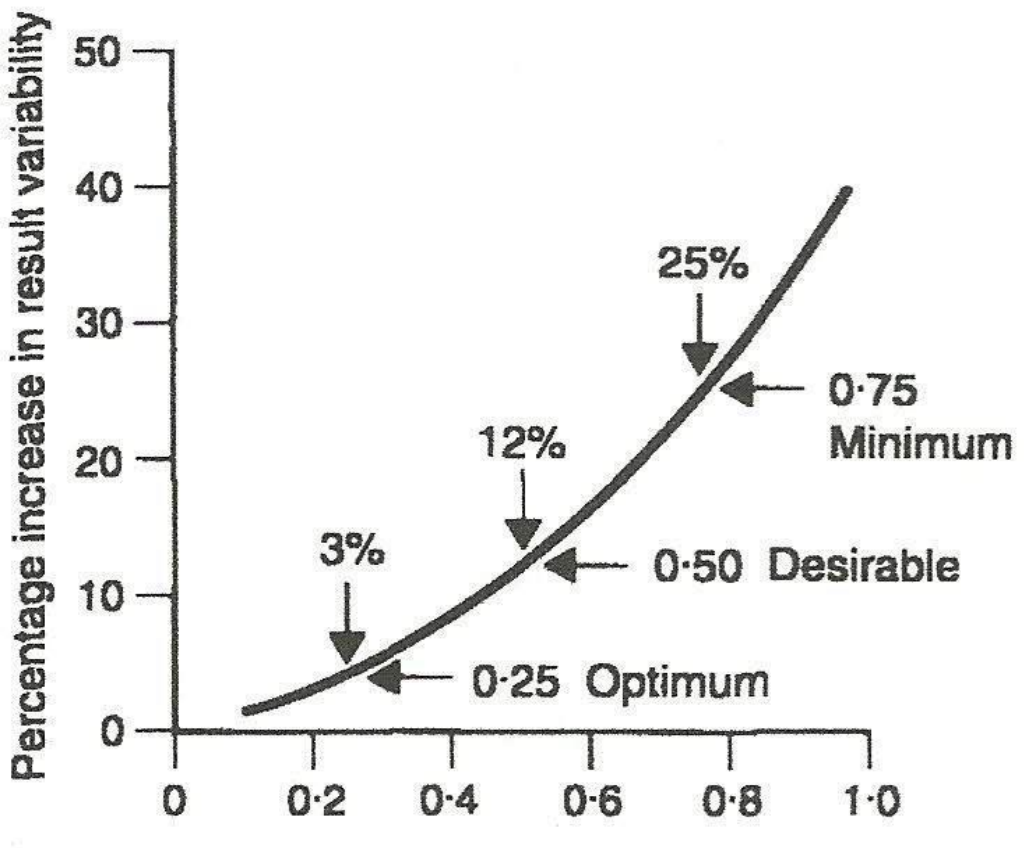
The clinical context requires the minimum total variability of test results (V_{TOT})

The contribution of CV_I to V_{TOT} cannot be changed

So that, the contribution of analytical variability to V_{TOT} should be minimized

If the intra-individual BV is high, the analytical requirements are relatively low. If, on the other hand, the intra-individual BV is low, it increases the necessity to reduce the analytical part of the total variation.

The BV-based APS model aims to minimize analytical noise relative to the BV of the measurand by defining APS as some fraction of the expected BV



[Fraser CG et al., Ann Clin Biochem 1997;34:8]

TFG on Biological Variation Database (TFG-BVD)

Terms of Reference: To use a critical appraisal check list to evaluate literature on biological variation.

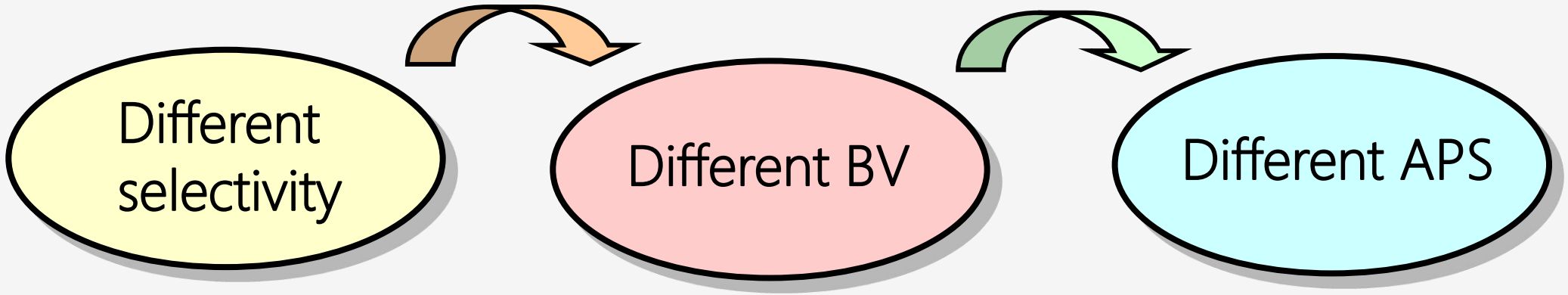
Deliverable: To generate a database on the EFLM website with essential information about the biological variation for different measurands as well as the evidence behind.

Current limitations of BV database

- Different selectivity of methods employed for some measurands in different studies not considered
- Influence of insufficient method sensitivity to detect measurands present in plasma at very low concentrations in all samples of all subjects enrolled in the BV protocols not considered
- The employed strategy based on the meta-analysis of available data may expose to flaws because including in the meta-analysis studies showing significant heterogeneity and low quality cannot provide an accurate information for APS derivation.
- Lacking of information on measurands for which APS based on BV should not be used but other models should be preferred

Assay selectivity is an important BV qualifier

If the used methodology has different selectivity for the measurand, one can expect that also the BV, a property closely associated with the characteristics of the measurand itself, significantly changes. And, if the BV changes, the APS derived from it may be different.

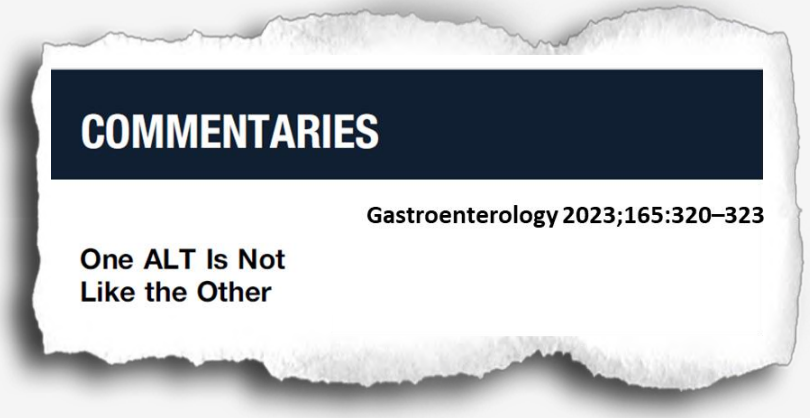


ALANINE AMINOTRANSFERASE (ALT)

Carobene A, Røraas T, Sølvik UØ, et al.; European Biological Variation Study of the EFLM Working Group on Biological Variation. Biological variation estimates obtained from 91 healthy study participants for 9 enzymes in serum. Clin Chem. 2017;63(6):1141

Pineda-Tenor D, Laserna-Mendieta EJ, Timón-Zapata J, et al. Biological variation and reference change values of common clinical chemistry and haematologic laboratory analytes in the elderly population. Clin Chem Lab Med. 2013;51(4):851

with pyridoxal-5-phosphate


COMMENTARIES

Gastroenterology 2023;165:320–323

One ALT Is Not Like the Other

Ma L, Zhang B, Luo L, et al. Biological variation estimates obtained from Chinese subjects for 32 biochemical measurands in serum. Clin Chem Lab Med. 2022;60(10):1648

Hölzel WG. Intra-individual variation of some analytes in serum of patients with insulin-dependent diabetes mellitus. Clin Chem. 1987;33(1):57

Wang S, Zhao M, Su Z, Mu R. Annual biological variation and personalized reference intervals of clinical chemistry and hematology analytes. Clin Chem Lab Med. 2022;60(4):606

without pyridoxal-5-phosphate

CV_I meta-analysis of 5 studies = 11.4% vs. CV_I from Carobene et al. = 9.3%

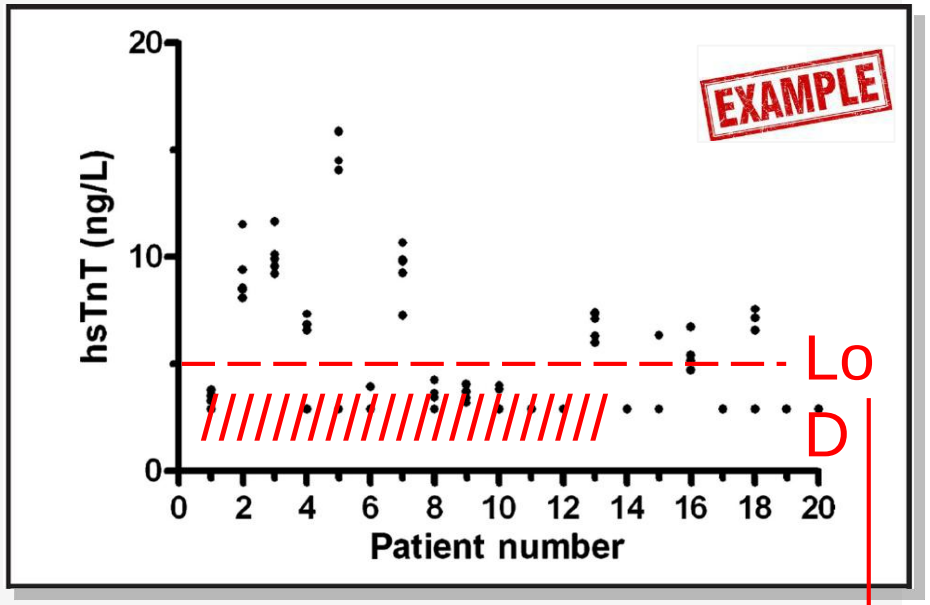
APS for standard measurement uncertainty from meta-analysis = 5.70% vs. APS from Carobene et al. = 4.65%

Current limitations of BV database

- Different selectivity of methods employed for some measurands in different studies not considered
- Influence of insufficient method sensitivity to detect measurands present in plasma at very low concentrations in all samples of all subjects enrolled in the BV protocols not considered
- The employed strategy based on the meta-analysis of available data may expose to flaws because including in the meta-analysis studies showing significant heterogeneity and low quality cannot provide an accurate information for APS derivation.
- Lacking of information on measurands for which APS based on BV should not be used but other models should be preferred

Measurands present in plasma at very low (often undetectable) concentrations

Most studies which have tried to assess BV of cardiac troponins provide data which are unworkable as a significant number of results for selected individuals were <LoD, even when highly sensitive assays were employed, preventing accurate measurement of random physiological fluctuations around the homeostatic set-point of this measurand.



[Frankenstein L et al., Clin Chem 2011;57:1068]

Only studies utilizing assays able to measure troponin *in all samples of all enrolled subjects* will deliver robust information on BV of this measurand without any result selection bias.

PLEASE NOTE

Limit of Detection (LoD): the lowest amount of troponin in a biological sample that can be detected by the assay

Current limitations of BV database

- Different selectivity of methods employed for some measurands in different studies not considered
- Influence of insufficient method sensitivity to detect measurands present in plasma at very low concentrations in all samples of all subjects enrolled in the BV protocols not considered
- The employed strategy based on the meta-analysis of available data may expose to flaws because including in the meta-analysis studies showing significant heterogeneity and low quality cannot provide an accurate information for APS derivation.
- Lacking of information on measurands for which APS based on BV should not be used but other models should be preferred

‘Combine *vs.* do-not-combine’ in EFLM database

In a meta-analysis the quality and strength of the deductions are only as strong as the quality of the studies included in the analysis

ON MY MIND

Are Meta-Analyses a Form of Medical Fake News?

Milton Packer,
Cardiologist in Dallas, TX

‘Combine *vs.* do-not-combine’ in EFLM database

If BV studies of high quality are available for a given measurand, it is probably better to derive APS estimates from such studies alone rather than from meta-analysis results

The latter approach has been used in the APERTURE study*, a project for establishing Analytical Performance Specifications for Measurement Uncertainty, where the BV publications with the highest quality rate were retrieved and directly used to derive APS for MU

* Braga F, Panteghini M. Clin Chem Lab Med 2021;59:1362
Braga F et al. Clin Chem Lab Med 2023;61:213
Borrillo F et al. J Appl Lab Med 2023;8:420

CV_I and derived desirable APS for standard measurement uncertainty (MU) on clinical samples for measurands having characteristics for being allocated to BV-based model, as derived in the APERTURE project and on the EFLM database

Measurand	APERTURE			EFLM database		
	CV_I	**Desirable APS**	**Source of BV data (highest quality rate study)**	**CV_I**	**Desirable APS**	**Source of BV data**
LDH	5.20%	2.60%	Carobene A et al. [2017]	4.40%	2.20%	Meta-analysis of 2 studies
IgG	4.40%	2.20%	Ford RP et al. [1988]	3.50%	1.75%	Meta-analysis of 2 studies
IgA	5.00%	2.50%	Ford RP et al. [1988]	7.50%	3.75%	Meta-analysis of 2 studies
Homocysteine	7.04%	3.52%	Garg UC et al. [1997]	6.10%	3.05%	Meta-analysis of 3 studies
D-dimer	21.2%	10.6%	Ercan Ş et al. [2021]	25.2%	12.6%	Meta-analysis of 5 studies

Current limitations of BV database

- Different selectivity of methods employed for some measurands in different studies not considered
- Influence of insufficient method sensitivity to detect measurands present in plasma at very low concentrations in all samples of all subjects enrolled in the BV protocols not considered
- The employed strategy based on the meta-analysis of available data may expose to flaws because including in the meta-analysis studies showing significant heterogeneity and low quality cannot provide an accurate information for APS derivation.
- Lacking of information on measurands for which APS based on BV should not be used but other models should be preferred

EUROPEAN FEDERATION OF CLINICAL CHEMISTRY AND LABORATORY MEDICINE

EFLM Biological Variation Database

Search for analyte | Search

Meta - Analysis | List of all BV Estimates | Measurands

Examples of APS for MU (desirable and minimum) of measurands allocated to BV model

Biological variation model		
Serum sodium	0.27	0.40
Serum potassium	1.96	2.94
Serum chloride	0.49	0.74
Serum total carbon dioxide	2.10	3.15
Serum total calcium	0.91	1.36
Serum inorganic phosphate	3.84	5.75
Serum magnesium	1.44	2.16
Serum creatinine	2.20	3.30
Serum urea	7.05	10.6
Serum urate	4.16	6.24
Plasma lactate	13.6	20.4
Serum total bilirubin	10.5	15.7
Serum alkaline phosphatase	2.65	3.98
Serum aspartate aminotransferase	4.75	7.13
Serum γ-glutamyltransferase	4.45	6.68
Serum lactate dehydrogenase	2.60	3.90
Serum cholinesterase	2.10	3.15
Serum total proteins	1.30	1.95
Serum IgG	2.20	3.30
Serum IgA	2.50	3.75
Serum IgM	2.95	4.43
Serum prostate-specific antigen	3.40	5.10
Plasma homocysteine	3.52	5.27
Red blood cells	1.55	2.33
White blood cells	5.65	8.48
Serum free T3	2.35	3.53
Serum free T4	2.80	4.20
Serum parathyroid hormone	7.85	11.8

Temporarily belonging to biological variation model[c]		
Serum albumin	1.25	1.88
Plasma D-dimer	10.6	15.9
Blood platelets	4.85	7.28
Serum alanine aminotransferase	4.65	6.98
Serum creatine kinase	7.25	10.9
Serum pancreatic lipase	3.85	5.78
Serum pancreatic amylase	3.15	4.73

[Panteghini M. Clin Chem Lab Med 2024;62:1497]

Model 3: Based on the state of the art

- Applied when a measurand has neither a central diagnostic role nor strict homeostatic control.
- This model can be temporarily used also for those measurands still waiting for the definition of outcome-based APS or for which the BV-based model should not be used because a strict homeostatic control is lacking.

Serum CRP: a measurand that does not have the biological and clinical characteristics to be allocated to models 1 and 2

Uses of CRP testi

- First-level test for inflammatory ap s disease
- Assessment of the extent of th atory disease activity
- Differential diagnosis: ba ral infection, Crohn's disease vs. ulcerative colitis, R
- Monitoring of th to anti-inflammatory or antibiotic therapy
- Referen nterpretation of variations of other plasma prot erritin)

Hov levation of CRP is not diagnostic of any one specific disease as it urs in many diseases involving tissue damage or inflammation.

Model 1 for defining APS for CRP cannot be recommended

CRP is also a biologically challenging analyte

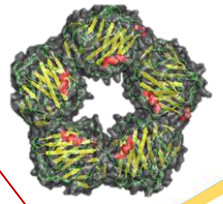
EFLM

EUROPEAN FEDERATION OF CLINICAL C
AND LABORATORY MEDIC

EFLM Biological V abase

Matrix	BV	lower CI limit	higher CI limit
Serum/pl	34.1	29.4	74.4

Model 2 for defining APS for CRP is also unsuitable



An elevation of CRP is not diagnostic of any one specific disease

Has the measurand a central role in a specific disease?

Yes

Do valid outcome data exist?

Yes

Assign to **outcome** model

No

Temporarily

NO

Produce **outcome** data

CRP is a biologically challenging analyte

Has the measurand a steady state?

Yes

Do valid biological variation data exist?

Yes

Assign to **biological variation** model

No

Produce **biological variation** data

NO

Temporarily

Assign to **state-of-the-art** model

The measurand does not have the characteristics to be allocated to models 1 and 2

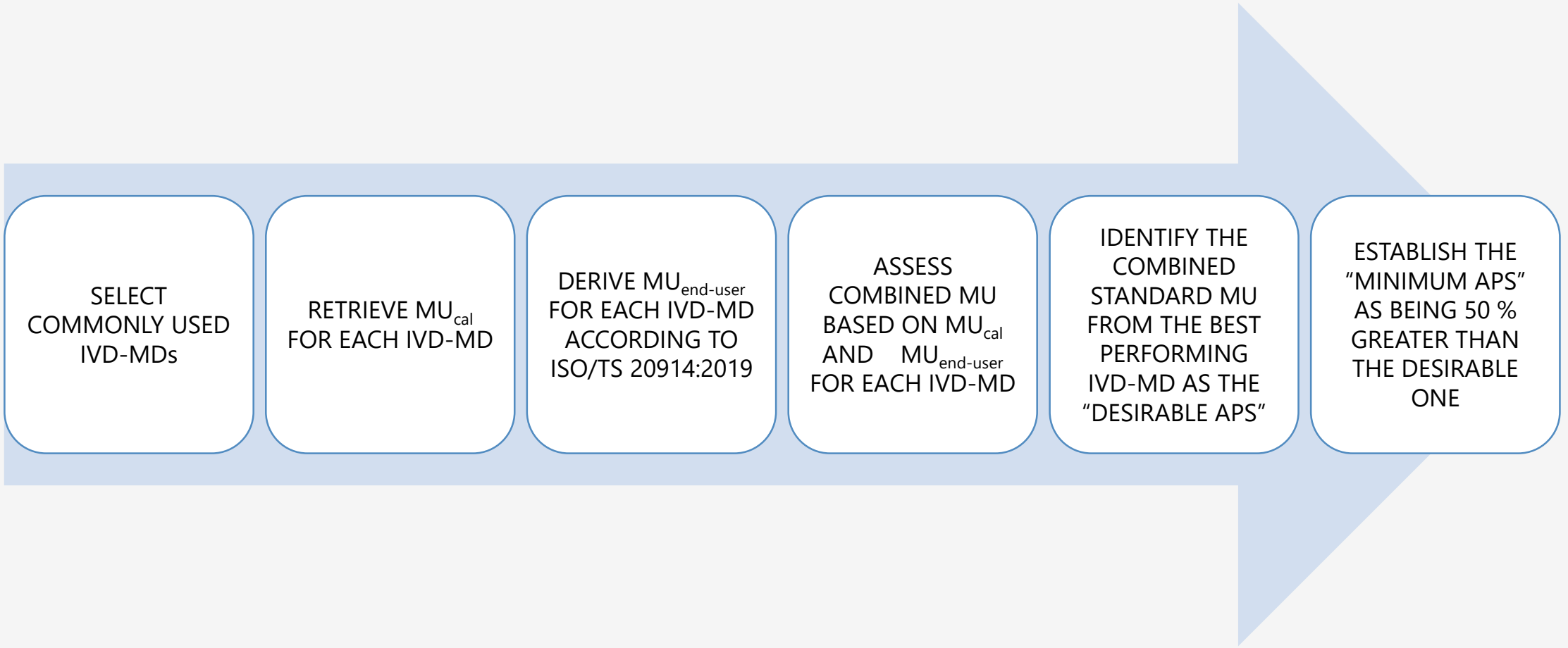
Model 3: Based on the state of the art – Problems with the definition: is the SA related by the best achieved or the current observed quality?

State-of-the-art has been defined as:

- "th
  ac
- "th
  EQ
- "the mean performance declared for that test by the most relevant manufacturers".

The SA definition for deriving APS should be related to an aspirational approach that links this definition to the best quality available.

Proposed approach to define the state of the art of measurement uncertainty as the highest level of performance technically achievable using the ISO/TS 20914 guidance for the MU estimate



[Borrillo F & Panteghini M. Clin Chem Lab Med 2024;62:1490]

Measurand	APS for standard MU		References
	Desirable	Minimum	
C-reactive protein	3.76%	5.64%	Braga F & Panteghini M [2020]
Intact human chorionic gonadotropin	4.55%	6.83%	Panteghini M [2024]
Temporarily belonging to state-of-the-art model			
Ferritin	4.31%	6.47%	Rovegno L et al. [2024]
Thyroid stimulating hormone	2.89%	4.34%	Borrillo F et al. [2023]
Urine total protein	4.97%	7.46%	Borrillo F & Panteghini M [2024]

- ✓ The myth of state-of-the-art as a ‘rescue’ model when APS correctly obtained with other more appropriate models for a certain measurand appear too stringent should be dismantled.
- ✓ Using APS derived from the correct allocation of measurands in different models has been shown helpful in identifying measurands that need analytical improvement for their clinical use.

Hybrid model for drugs

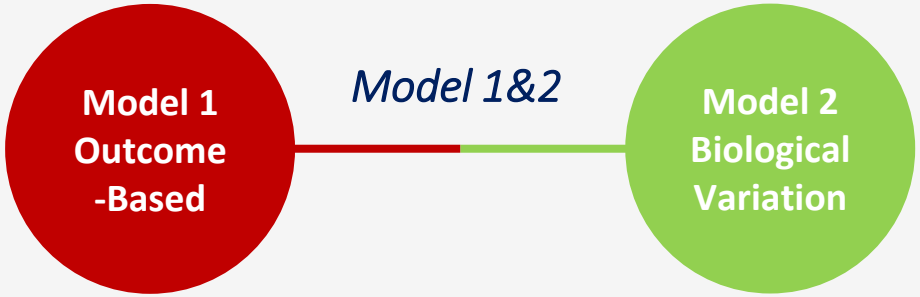
Drugs need a dedicated approach when deriving APS, based on fundamental pharmacokinetic theory and average elimination half-life of the drug

$$\text{Desirable MU} \leq 1/4\ [(2^{T/t} - 1)/(2^{T/t} + 1)] \times 100\%$$

where T is the time interval between doses and t is the average elimination half-life of drug

[Adapted from Fraser C, Clin Chem 1987;33:387]

Although concentrations of drugs do not fluctuate randomly around a homeostatic set point, this approach has a relationship with biological knowledge. On the other hand, TDM is linked to the patient outcome in defining the levels of drug which are potentially toxic or when the treatment can be ineffective. Accordingly, a *hybrid model between the models 1 and 2* has been proposed for drugs.



[Braga F et al. Clin Chem Lab Med 2023;61:213]

APS for MU in therapeutic monitoring of immunosuppressive drugs

Drug	Frequency of administration, h	Terminal half-life, h	APS for u_{result}, %	
			Desirable	Minimum
Cyclosporine	12	8.4	11.5	17.2
Everolimus	12	25.6	4.03	6.04
Sirolimus	24	60.0	3.44	5.17
Tacrolimus, immediate release	12	12.1	8.27	12.4
Tacrolimus, prolonged release	24	36.0	5.68	8.51
Tacrolimus, extended-release tablets (Envarsus XR®)	24	30.0	6.76	10.1

[Cattaneo D & Panteghini M. Clin Chem Lab Med 2024;62:e81]

[Ceriotti F et al. Clin Chem Lab Med 2024;62:1470]

Ten years after Milan conference: how medical laboratories select and use APS? A national survey

The meaning of the term APS is still not well known. It is suspected that most of the “non repliers” in the proposed survey did not know the topic.

Even among laboratorians that are aware of the topic a significant group declared not being using APS. So that, quality specifications still remain largely unapplied.

This situation challenges us as laboratory professionals. If we think APS are useful, we need to put more efforts to educate laboratorians in how to use them and for what intended purposes.

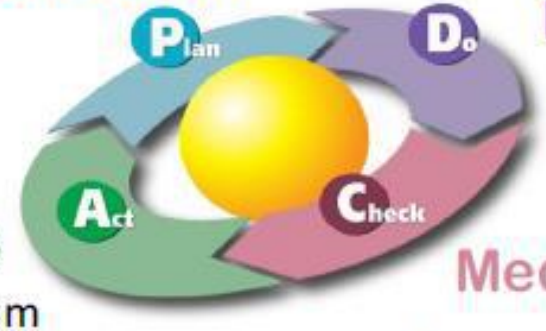
Roles and main actions expected from each stakeholder for contributing a quantum leap forward in the way of practicality of Milan consensus about APS

Endorse and adopt APS defined according to the Milan principles, and complete the allocation of laboratory measurands to appropriate APS models

Laboratory profession

Give more focus to the minority of tests that need urgent improvement and operate to reduce as much as possible the measurement uncertainty of calibrators, especially when APS are stringent

IVD manufacturers



All stakeholders

Work together for optimum patient care by applying suitable APS

Medical laboratories

Supervise the IVD-MD quality and their own performance using appropriate IQC and EQA approaches, which should include objectively derived APS

[Panteghini M. Clin Chem Lab Med 2024;62:1455]


Thank you for listening
QUALITY
HELP
THE DOCTOR
IS IN
SCHULZ