# CIRME



Università degli Studi di Milano

Centre for Metrological Traceability in Laboratory Medicine (CIRME)

Director: Prof. Mauro Panteghini

site: http://users.unimi.it/cirme

9<sup>th</sup> International Scientific Meeting STRUCTURING EQAS FOR MEETING METROLOGICAL CRITERIA: READY FOR PRIME TIME

> MILANO, ITALY November 27<sup>th</sup>, 2015

# Performance specifications in EQAS Mauro Panteghini

		Table 3.	Evaluatio	on capabilitio	es of PT	/EQA re	elated to s	cheme design		
							Evaluation	capability		
				Ac	curacy					
				Individua	al laborat	ory			Standardi harmon	
	Sample characteristics			Relative ticipant		Repro	oducibility	Measuremer calibration		
Category	Commutable	Value assigned with RMP <sup>a</sup> or CRM	Replicate samples in survey	Absolute vs RMP or CRM	Overall	Peer group	Individual laboratory intralab CV	Measurement procedure interlab CV	Absolute vs RMP or CRM	Relative to participant results
1	Yes	Yes	Yes	Х	Х	Х	Х	Х	Х	Х
2	Yes	Yes	No	Х	Х	Х		Х	Х	Х
3	Yes	No	Yes		Х	Х	Х	Х		Х
4			No		Х	Х	V	X		Х
5 6			Yes No			X X	Х	X X		

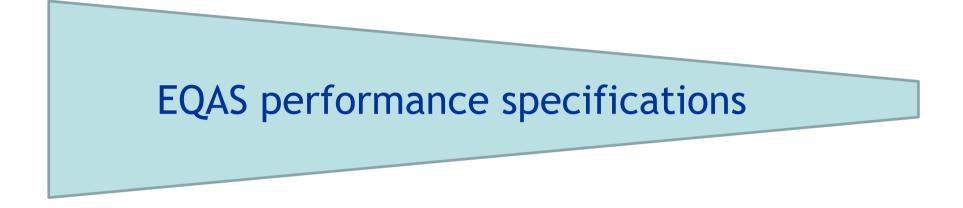




# EQAS performance specifications

- Currently wide variation in practice
- Range between very "tight" and very "loose"
- May be based on:
  - Clinical
  - Biological variation
  - State of the art (different definitions)
  - Statistical
  - Regulatory
  - Combination of models (e.g. state of the art and BV)
  - Other (e.g. professional recommendations)





## (looser) e.g. regulatory All labs pass

### (tighter)

e.g. biological A portion of labs fail



### **GENERAL PAPER**

### Comparison of evaluation procedures used by European external quality assessment scheme organizers for haemoglobin concentration and leukocyte concentration

Table 1 Criteria used for acceptable performance for haemoglobin concentration in blood and leukocyte concentration (deviation from the target value)

Table 2 Percentages of unsatisfactory results reported by the participating EQAS organizers for a fixed set of 262 results of haemoglobin concentration in blood and leukocyte concentration

	Scheme	Haemoglobin Leukocyte		haemoglobin concentration in blood and leukocyte concentration				
		concentration	concentration	Scheme	Haemoglobin	Leukocyte		
	Belgium	$\pm 2s$	$\pm 2s$		concentration	concentration		
	France	$\pm 2s$	$\pm 2s$	Belgium	6.9	7.3		
	Spain (two organizers)	$\pm 2s$	$\pm 2s$	Croatia	14.9	15.3		
	Croatia	$\pm 1s$	$\pm 1s$	Finland	1.5	3.1		
	Germany	$\pm 6\%$	$\pm 18\%$	France	5.4	4.6		
	Finland	$\pm 5\%$	$\pm 10\%$	Hungary: consensus mean	13.5	19.8		
	Hungary: consensus	$\pm 3\%$	$\pm 6\%$	method				
	mean	$\pm 5\%$	±15%	Russia	15.6	19.8		
	Hungary: target value set by reference labs or	土3%	±13%	Spain 1	7.6	4.6		
	manufacturers			Spain 2	3.1	2.3		
	Russia	$\pm 1.64s$	$\pm 1.64s$	Switzerland: QUALAB	0.4	0		
			(official for licensing)					
CIRI	Switzerland: QUALAB (official for licensing)	$\pm 9\%$	±25%	Switzerland: CSCQ (scientific approach)	0.8	2.0		
	Switzerland: CSCQ $\pm 3\%$ $\pm 8\%$ (scientific approach)		$\pm 8\%$	New York State, USA	0.8	2.3		
Università de	New York State, USA	±7%	$\pm 15\%$					

UNIVERSITÀ D DI MILANU Requirements for the applicability of EQAS results in the evaluation of the performance of participating laboratories in terms of traceability of their measurements

Feature	Aim
EQAS materials value-assigned with reference procedures by an accredited ref. laboratory	To check traceability of commercial system to reference systems
Proved commutability of EQAS materials	To allow transferability of participating laboratory performance to the measurement of patient samples
Definition and use of the clinically allowable ( measurement error	To verify the suitability of laboratory measurements in clinical setting
Università degli Studi	Panteghini M, CCLM 2010;48:7 Infusino I et al., CCLM 2010;48:301 Braga F & Panteghini M. CCLM 2013;51:1719

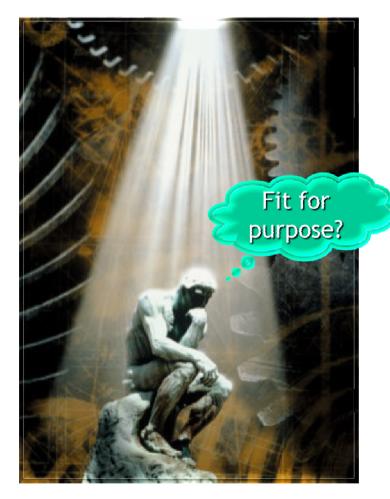
Braga F & Panteghini M, Clin Chim Acta 2014;432:55

Analytical performance specifications: definition

Criteria that specify (in numerical terms) the quality required for analytical performance in order to deliver laboratory test information that would satisfy clinical needs for improving health outcomes.



# The Essential Question...



"What amount of medical harm due to analytical error is it OK to let go undetected?"



# Steps of the process and different responsibilities in implementing traceability of patient results and defining their uncertainty

Profession (e.g., JCTLM, IFCC, EFLM): Define analytical objectives: reference measurement systems (traceability chain) and associated clinically acceptable uncertainty (fit for purpose)

Diagnostic manufacturers:

Implement suitable analytical systems (platform, reagents, calibrators, controls) fulfilling the above established goals

End users (clinical laboratories):

Survey assay and laboratory performance through IQC and EQA redesigned to meet metrological criteria



Adapted from Panteghini M, Clin Chem Lab Med 2010;48:7

# **QUALITY SPECIFICATIONS IN** STRATEGIES TO SET GLOBAL LABORATORY MEDICINE



WORLD HEALTH ORGANIZATION ORGANISATION MONDIALE DE LA SANTE



Pure and Applied Chemistry International Union of









# Stockholm April 24-26, 1999 Karolinska Institutet Nobelforum,



UNIVERSITÀ DEGLI STUDI DI MILANO

# 1999 Stockholm Consensus revised in Milan 2014

Although the essence of the hierarchy established in Stockholm was supported, new perspectives have been forwarded prompting simplification and explanatory additions.

The most innovative aspect of the new consensus is that it is recognized that some models are better suited for certain measurands than for others; the attention is therefore primarily directed towards the measurand and its biological and clinical characteristics.





	EFLM		DE GRUYTER Clin Chem Lab Med 2015; aop
	A TRANSPORTER AND A TRANSPORT	Citate State	Sverre Sandberg*, Callum G. Fraser, Andrea Rita Horvath, Rob Jansen, Graham Jones, Wytze Oosterhuis, Per Hyltoft Petersen, Heinz Schimmel, Ken Sikaris and Mauro Panteghini
	Defining conference		Defining analytical performance specifications: Consensus Statement from the 1st Strategic
	performance goals 15 years after the		Conference of the European Federation of Clinical Chemistry and Laboratory Medicine
	PIOCKholm Conference <sup>8^</sup> CIRME International Scientific Meeting	ence	Model 1: Based on the effect of analytical performance on clinical outcomes
	Milan (IT) 24-25 November 2014	Program	a. Done by direct outcome studies – investigating the impact of analytical performance of the test on clinical outcomes;
	G F F F A L L RGBT RATION F FE EUR 200.00 (VM 22% incuded) The mg laterdow is incuded: P - Order branch Austra butk an indouted in the programma P - Order predicipation C - Might Brothon C - Might	G E R E R.A. L. I. N F O R. M.A.T. J O N VIBUE Androit Encodes Androit Encode Androit Androit Androit Context Androit Androit Androit Androit Encodes Androit Encodes Androit Encodes Androit Encodes Androit Androit Androit Encodes Androit Androit Androit Encodes Androit Encodes Androit Encodes Androit Encodes Androit Androit Androit Encodes Androit Encodes Androit Encodes Androit Encodes Androit Encodes Androit	b. Done by indirect outcome studies – investigating the impact of analytical performance of the test on clini- cal classifications or decisions and thereby on the probability of patient outcomes, e.g., by simulation or decision analysis.
	Ib make your majoration, pieze access the following latt. May Insymmetry commercial many access that way OF PICIA 1, LAN CLARGE The official imprage of the conference is English. CREAM DAY BEOR ELARIAT	<ul> <li>congress which. To book your soon please when to be base indicated host reservation system.</li> <li>colo Atahonai Bascute (conference annua) high Investitation (andionalise)</li> <li>colo MAR(pi Host) (atahonalise) the congress annual the assessment that has a second any structure.</li> </ul>	Model 2: Based on components of biological variation of the measurand.
C	ert 917 ert 917 ğınuconyr	Cohesis ACMisso (500 meter from the couples area) ing Anne analotic anthatakin relativishic conditioned colorisation from them the couples areas) ing Anne Anneling failed control is the time and uncond from 4 support for the kind and uncond from 4 support for the kind and uncond from 4 support	Model 3: Based on state of the art of the measurement (i.e., the highest level of analytical performance technically achievable).
	Abbout BIOHAD		

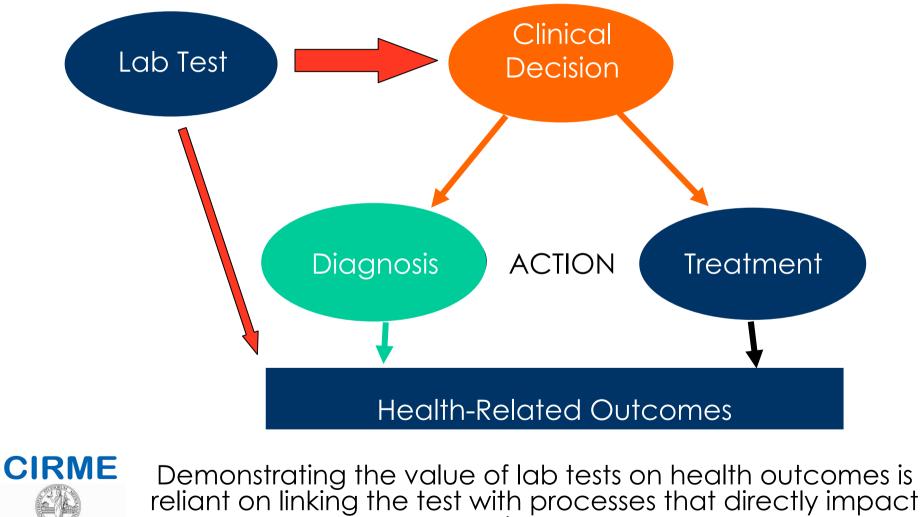
# Model 1. Based on the effect of analytical performance on clinical outcomes

- Advantage: to address the influence of analytical performance on clinical outcomes that are relevant to patients and society.
- Disadvantage: it is only useful for examinations where the links between the test, clinical decisionmaking and clinical outcomes are straightforward and strong. Furthermore, it may be influenced by the current measurement quality and results may vary according to the actual test method used, the investigated population and health care setting.





# Challenge: Connecting Laboratory Testing to Outcomes



Università degli Studi DI MILANO

outcomes

# Performance specifications based on clinical needs defined in terms of allowable misclassification rates

**Table.** Recommended analytical performance goals for cardiac troponin measurement for definition of the limit of quantitation of assays.

Quality level		Imprecision goal (as CV)		Bias goal <sup>a</sup>
	Outcome-based	Biological variability <sup>a</sup>	Expert opinion	
Minimum Desirable Optimum	<13% <sup>b</sup> <10% <sup>c</sup> <6% <sup>d</sup>	<7.3% <4.9% <2.4%	<20% <10%	±21.6 % ±14.4 % ±7.2 %

<sup>a</sup> Calculated according to Fraser CG, Hyltoft Petersen P, Libeer JC, Ricos C. Proposal for setting generally applicable quality goals solely based on biology. Ann Clin Biochem 1997;34:8-12. <sup>b</sup> Assuming a diagnostic misclassification of 1.8%, <sup>c</sup> 1.0%, and <sup>d</sup> 0.5%.





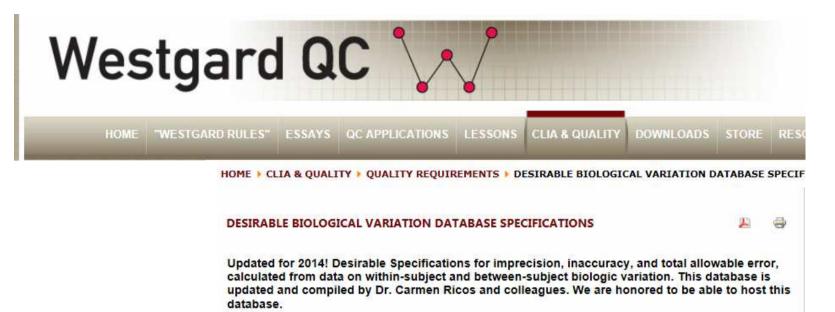
Panteghini M, AACB Troponin Monograph 2012

# Model 2. Based on components of biological variation of the measurand

- Advantage: it can be applied to most measurands for which a "steady state" biologic model can be established.
- Disadvantage: need to carefully assess the relevance of biological variation data.







DE GRUYTER

Clin Chem Lab Med 2015; 53(2): 299-305

Carmen Perich, Joana Minchinela, Carmen Ricós\*, Pilar Fernández-Calle, Virtudes Alvarez, María Vicenta Doménech, Margarita Simón, Carmen Biosca, Beatriz Boned, José Vicente García-Lario, Fernando Cava, Pilar Fernández-Fernández and Callum G. Fraser

# Biological variation database: structure and criteria used for generation and update



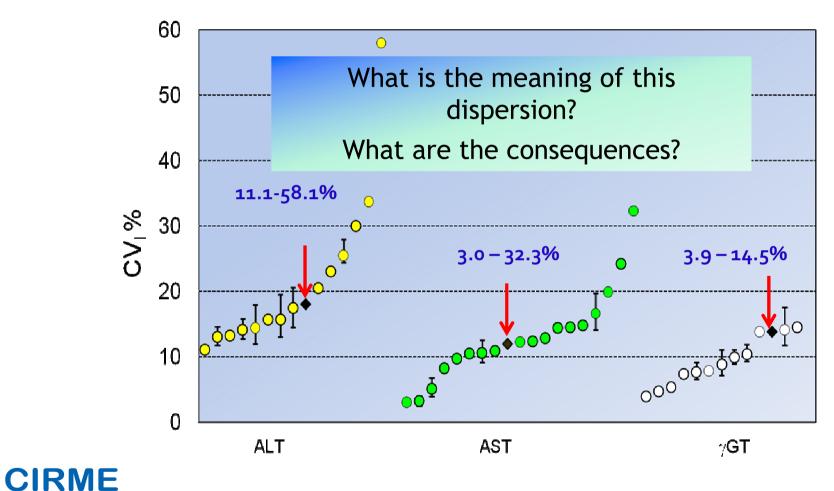
More than 240 articles More than 350 measurands



Università degli Studi di Milano Generation of estimates of  $CV_I$  and  $CV_G$  using the MEDIAN of all data compiled

### ALT, AST and $\gamma \rm GT$





The arrows show the values currently present in the Ricos' database

Università degli Studi di Milano

Carobene A et al., Clin Chem Lab Med 2013;51:1997

# Quantifying Biological Variation

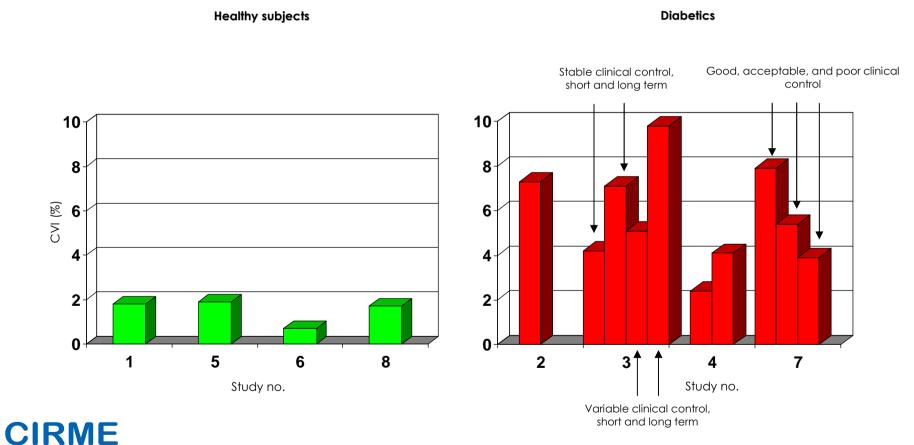
How do you do the experiment?

✓ Subjects
 ✓ Collect specimens
 ✓ Analyse specimens
 ✓ Analyse data

How many? Number? Frequency? Minimise analytical variation? Outliers? Statistics?



# Biological variation from patients Should they be used?



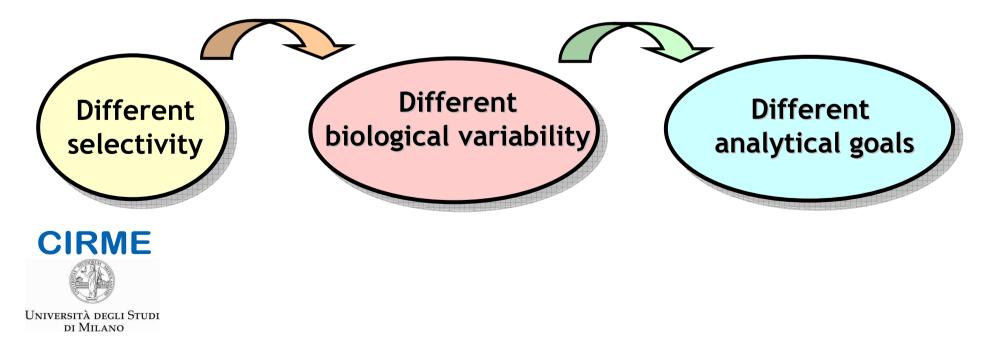
Intra-individual variation in pathology >> CV<sub>1</sub> of healthy individuals

Università degli Studi di Milano

Braga F et al, Chim Clin Acta 2010;411:1606

# Assay selectivity is an important biological variation qualifier

If the used methodology has different specificity for the measured analyte, one can expect that also the biological variability, a property closely associated with the characteristics of the analyte itself, significantly changes. And, if the biological variability changes, the analytical goals derived from it may be different.



8 journal homepage: www.elsevier.com/locate/clinchim Contents lists available at SciVerse ScienceDirect Clinica Chimica Acta Clinica Chimica Acta 413 (2012) 1179-1183 Invited critical review SEVIE

Biologic variability of C-reactive protein: Is the available information reliable?

Federica Braga \*, Mauro Panteghini

# Table 2

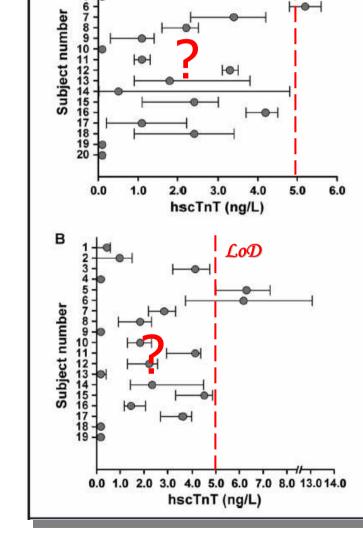
Summary of the characteristics of studies on biologic variability of C-reactive protein (CRP) evaluated in this systematic review

Summa	ry of the chara	summary of the characteristics of studies on biologic variability of	biologic variability of C-reactive pro	otein (CKP) evaluate	L-reactive protein (LKP) evaluated in this systematic review.		
Study	Assay	Recruitment of	Optimal study duration and	Appropriate	Optimal protocol of	Statistical test for Testing normal	Testing normal
no.	sensitivity	healthy subjects	sampling frequency	sample type	sample analysis	outliers	distribution of data
1	No	Yes	No	Yes	No	Yes	No
2	No	Yes	No	Yes	No	Yes	No
3a	No	Yes	No	No	NA	No	No
3b	No	No	No	No	NA	No	No
4	Yes	Yes	No	No	NA	Yes	No
5	Yes	e H	No	No	NA	No	No
9	Yes	Yes	Yes	Yes	Yes	Yes	No
2	Yes	No	Yes	Yes	Yes	No	Yes
8	Yes	Yes	No	Yes	No	No	No
6	Yes	р Н	No	No	No	Yes	No
10	Yes	No	No	No	NA	No	No
11	NA	No	No	Yes	No	Yes	No
NA, info	NA, information not available.	vailable.					

CIRME 

UNIVERSITÀ DEGLI STUDI di Milano

# Is available information on biological variability of troponin T reliable?



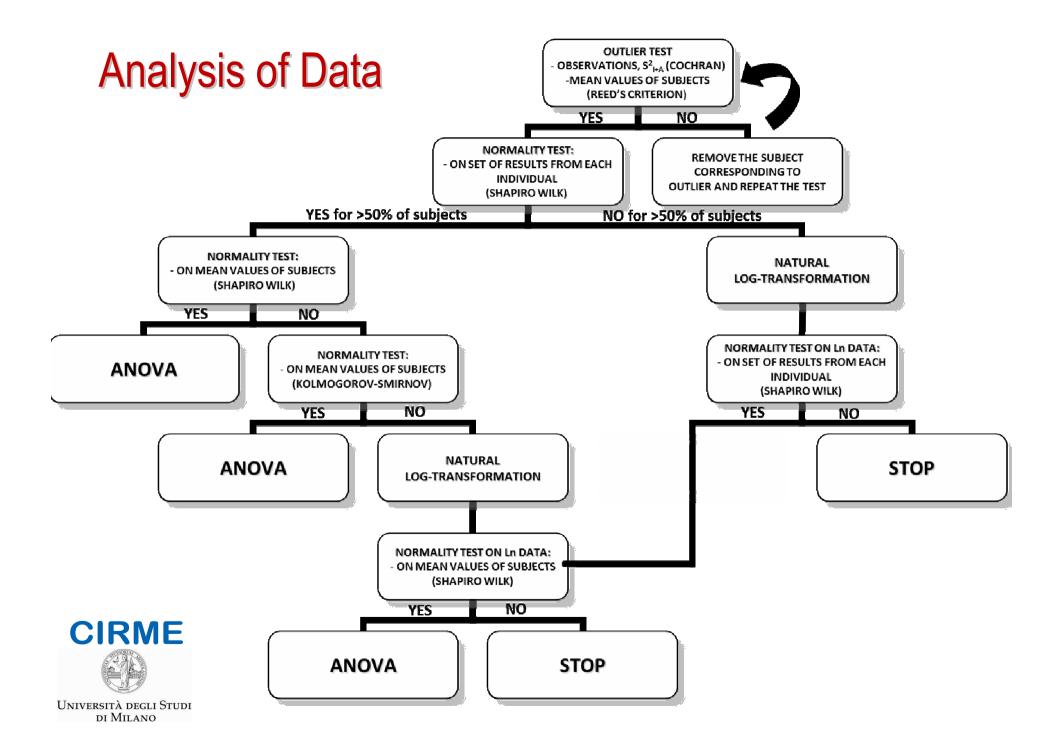
H-O-I

<u>Short-term</u>

<u>Long-term</u>



Vasile VC et al., Clin Chem 2010;56:1086



S and Biologic variability of C-reactive protein: Is the available information reliable? journal homepage: www.elsevier.com/locate/clinchim Contents lists available at SciVerse ScienceDirect Clinica Chimica Acta Clinica Chimica Acta 413 (2012) 1179-1183 Invited critical review

Table 2

Federica Braga \*, Mauro Panteghini

Summary	y of the chara	Summary of the characteristics of studies on biologic variability	1000	tein (CRP) evaluate	of C-reactive protein (CRP) evaluated in this systematic review.		
Study no.	Assay sensitivity	Recruitment of healthy subjects	Optimal study duration and sampling frequency	Appropriate sample type	Optimal protocol of sample analysis	Statistical test for outliers	Testing normal distribution of data
1	No	Yes	No	Yes	No	Yes	No
2	No	Yes	No	Yes	No	Yes	No
3a	No	Yes	No	No	NA	No	No
3b	No	No	No	No	NA	No	No
4	Yes	Yes	No	No	NA	Yes	No
5	Yes	т <sup>а</sup>	No	No	NA	No	No
9	Yes	Yes	Yes	Yes	Yes	Yes	No
7	Yes	No	Yes	Yes	Yes	No	Yes
8	Yes	Yes	No	Yes	No	No	No
6	Yes	<mark>م</mark> ۲	No	No	No	Yes	No
10	Yes	No	No	No	NA	No	No
11	NA	No	No	Yes	No	Yes	No

NA, information not available.



UNIVERSITÀ DEGLI STUDI DI MILANO



# Have data not normally distributed been appropriately transformed?



 This is a critical aspect!
 Studies using statistical parametric approach on data not normally distributed should not be considered! Otherwise we will continue to have CV >>33%! 2) When a not normal data distribution is present, a logtransformation of data is recommended, but this approach does not always solve the distribution problems!





	Analyte	Nun of	inder	Biolo Varia			Desira specif	ible ication	
		pap	ers	CVw	CVg	$\overline{\ }$	l(%)	B(%)	TE (%)
S-	C reactive protein	3		42.2	76.3	ノ	21.1	21.8	56.6
		$\square$	$\int$						

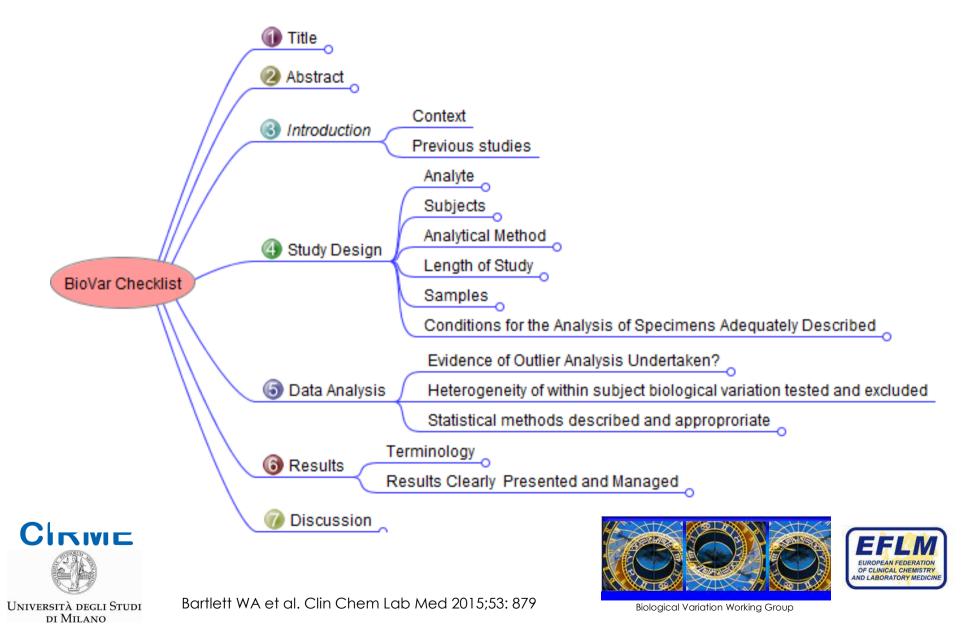


# Summary

- BV published data are of varying quality
- Safe application for deriving performance specifications requires prior critical appraisal
- Need for standards (i.e. a set of attributes to enable the data to be effectively transmitted and applied)



A checklist for critical appraisal of studies of biological variation



# Model 3. Based on state-of-the-art

This relates to the highest level of analytical performance technically achievable [but someone defines it as the analytical performance achieved by a certain percentage of laboratories]

Advantage: numbers are readily available.





# Problems with the state-of-the-art concept

- No scientific reasoning
- Often based on "old" data which may be outdated
- Lack of transparency
- Lack of neutrality (dependency on industry)
- No relationship between what is achievable and on what is needed clinically



Possible criteria for allocation of laboratory tests to different Milan models for performance specifications

- 1. The measurand has a central role in diagnosis and monitoring of a specific disease  $\Rightarrow$  outcome model
- 2. The measurand has a high homeostatic control  $\Rightarrow$  BV model
- 3. Neither central diagnostic role nor sufficient homeostatic control ⇒ state-of-the-art model

Università degli Studi di Milano

CIR

# Grading different quality levels

# IDEAL

## OPTIMUM STANDARD

# **DESIRABLE STANDARD**

# MINIMUM STANDARD UNACCEPTABLE



# Result interpretation in EQAS

- There should be two types of specifications, depending on the number of measurements of EQAS samples participants do.
- If participants measure in singlicate, a TE specifications is needed (when only a single measurement is performed, it is impossible to separate the different causes of measurement's deviation from the reference value).
- If EQAS ask for multiple measurements of the same sample, a bias specification should be used for judgement of an EQAS result, providing that the scheme is also able to independently estimate the random component of the measurement uncertainty of individual participants.

### Box 1 Factors influencing choice of External Quality Assessment (EQA) Scheme

- Accreditation status of provider. Preference should be given to schemes accredited to ISO 17043 or equivalent (eg, those still Clinical Pathology Accreditation (CPA) accredited within the UK). If a non-accredited provider is chosen, the reason(s) should be clearly documented. Under International Laboratory Accreditation Cooperation (ILAC),<sup>11</sup> accreditation bodies should support the use of appropriate proficiency testing programmes which meet the essential requirements of ISO/IEC 17043, where applicable.
- Appropriateness of distribution frequency. Distributions should be at a frequency sufficient to identify performance issues in a timely manner. For core tests, this probably equates to at least monthly distributions.
- Range and number of EQA samples. Samples within the distribution cycle should cover an appropriate range of values for each analyte to verify performance across clinically relevant concentrations. Each cycle should supply sufficient samples to provide evidence of reproducibility; 3–4 samples in each distribution would probably fulfil this requirement. Samples should be 'blinded' to participants in relation to expected results.
- Scheme management and development. The scheme should be designed and overseen by appropriately competent professionals (clinical, technical and statistical). The scheme should also have an independent medical and scientific committee.<sup>12</sup>
- Poor performance issues. Mechanisms should be in place for reporting of poor performance to the appropriate regulatory/ oversight body.
- Variety of sample provided. 'Challenging' samples should be included in selected distributions.
- Education. Educational input should be provided.
- Manufacturers. Participation of the EQA provider in postmarketing vigilance of in vitro diagnostics.<sup>12</sup>
- Materials. EQA providers should demonstrate use of commutable materials.<sup>13</sup>

James D et al., J Clin Pathol 2014;67:651



# **EQAS** categorization

							Evaluation	capability
				Ac	curacy			
				Individua	al laborat	ory		
	Sample	e characteris	tics		Relative ticipant		Repro	ducibility
Category	Commutable	Value assigned with RMP <sup>a</sup> or CRM	Replicate samples in survey	Absolute vs RMP or CRM	Overall	Peer group	Individual laboratory intralab CV	Measurement procedure interlab CV
1 2	Yes Yes	Yes Yes	Yes No	X X	X X	X X	Х	X X

Category 1A  $\rightarrow$  Milan model 1 or 2 as basis for PS Category 1B  $\rightarrow$  Other models





# The role of the Profession: "check"

1. Availability and quality of information about IVD metrological traceability and uncertainty

# 2. Daily surveillance of IVD system traceability

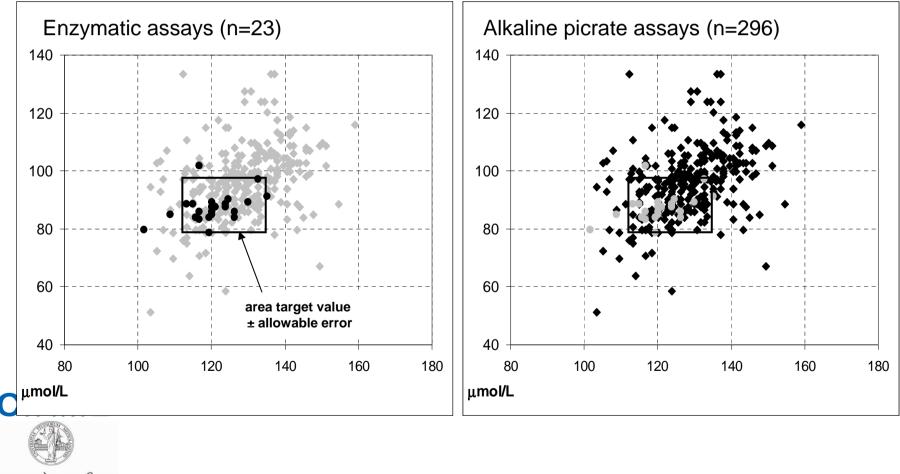
Verification of the consistency of declared performance during routine operations performed in accordance with the manufacturer's instructions

Participation to appropriately structured EQAS ("meeting metrological criteria")

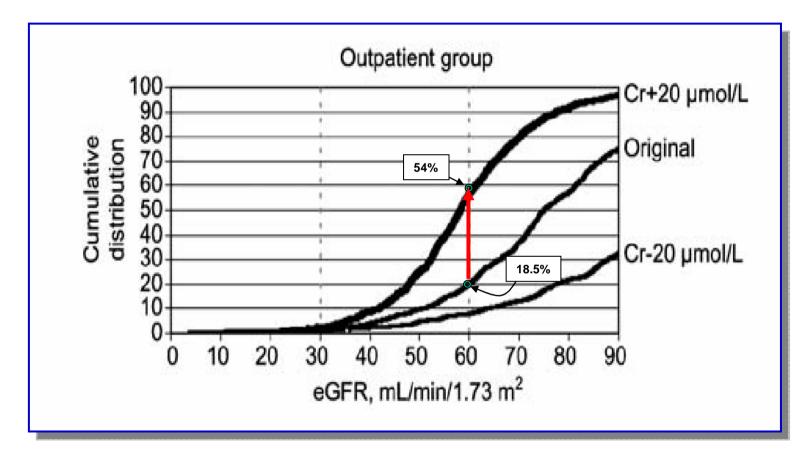




EQAS materials with physiologic (88.4 μmol/L) and borderline (123.8 μmol/L) creatinine concentrations vs. the desirable goal.
 The vast majority (87%) of laboratories using systems employing enzymatic assays were able to fulfill the desirable performance, while only one third of laboratories using picrate-based systems were able to meet the target.

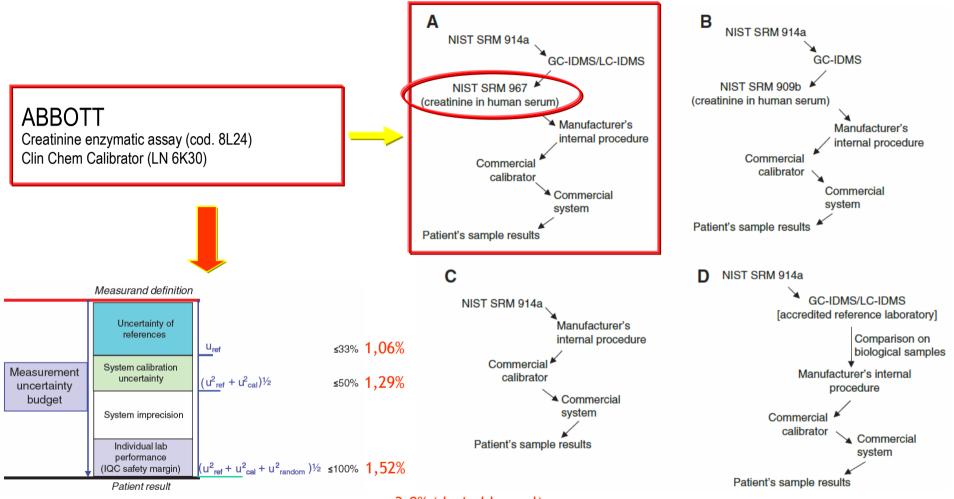


# Effect of analytic bias in creatinine on the distribution of estimated GFR values



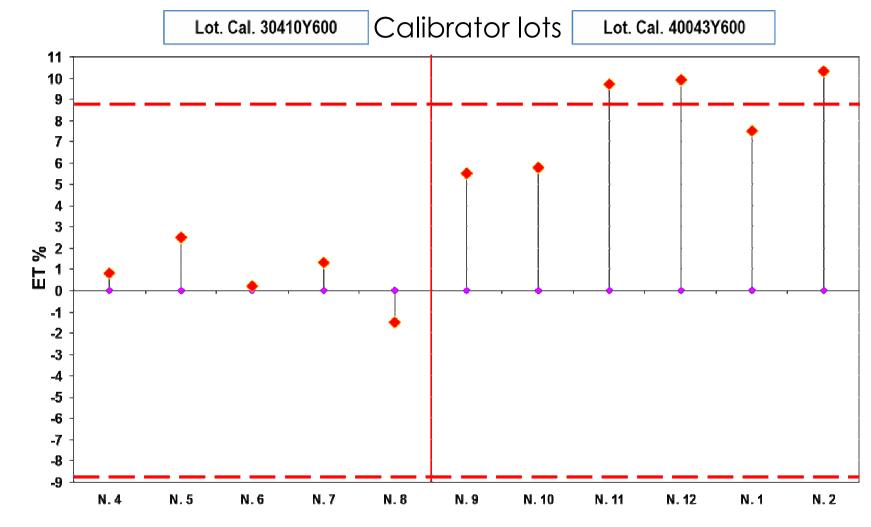


Klee GG et al., Clin Chem Lab Med 2007;45:737



vs. 3,0% (desirable goal)







### Pasqualetti S, Infusino I, Carnevale A, Szőke D, Panteghini M.

Clinica Chimica Acta 450 (2015) 125-6



Contents lists available at ScienceDirect

### Clinica Chimica Acta

journal homepage: www.elsevier.com/locate/clinchim

### Letter to the editor

The calibrator value assignment protocol of the Abbott enzymatic creatinine assay is inadequate for ensuring suitable quality of serum measurements

### Table 1

Uncertainties for each contributing factor in determination of serum creatinine with Abbott enzymatic assay on Architect c16000 platform after calibration with two different lot of system calibrator. Data obtained by measurements of NIST SRM 967a reference material (certified value  $\pm$  expanded uncertainty: L1, 0.847 mg/dL  $\pm$  0.018 mg/dL and L2, 3.877 mg/dL  $\pm$  0.082 mg/dL).

	SRM 967a level 1	SRM 967a level 2
Multigent Clin Chem Calibrator lot no. 40043Y600		
Imprecision (u <sub>Rw</sub> )	0.47%	0.40%
Bias (u <sub>bias</sub> )	3.57%	7.05%
Relative combined standard uncertainty $[u_c = (u_{bias}^2 + u_{Rw}^2)^{0.5}]$	3.60%	7.06%
Expanded uncertainty ( $U = k \times u_c$ )	7.20%	14.12%
Multigent Clin Chem Calibrator lot no. 40496Y600		
Imprecision (u <sub>Rw</sub> )	0.53%	0.42%
Bias (u <sub>bias</sub> )	4.02%	1.71%
Relative combined standard uncertainty $[u_c = (u_{bias}^2 + u_{Rw}^2)^{0.5}]$	4.05%	1.76%
Expanded uncertainty ( $U = k \times u_c$ )	8.10%	3.52%

Note: For serum creatinine measurements on patient samples, the acceptable limits for combined uncertainty derived from its CVI are 3.0% (desiderable) and 4.5% (minimum quality level), respectively.



# Basis for performance specifications

PROVIDER	MODELS
RCPAQAP Australia	Combination of BV and state of the art
SKML The Netherlands	Combination of BV and state of the art
NOKLUS Norway	Fixed percentage limits and based on a combination of BV, state of the art and expert opinion
SEQC Spain	Combination of BV and statistical results
WEQAS UK	Combination of BV and state of the art
SEHH Spain	Statistical/state of the art/BV
CTCB France	z-score/state of the art/limits given by scientific societies or other/limits based on clinical impact







TFG on Performance Specifications for EQAS

- Apply Milan models to describe EQAS performance specifications
- Develop common performance specifications based on Milan models
- Focus on "type 1" EQAS (commutable materials, reference measurement for target, repeated samples)



- Definition and approval of reference measurement systems, possibly in their entirety;
- Implementation by IVD industry of traceability to such reference systems in a scientifically sound and transparent way;
- Definition by the profession of the clinically acceptable measurement uncertainty (error) for each of the analytes used in the clinical field;
- Adoption by EQAS providers of commutable materials and use of an evaluation approach exclusively based on trueness;
- Monitoring of the analytical performance of individual laboratories by the participation in EQAS that meet metrological criteria and application of clinically acceptable limits;
- Abandonment by users (and consequently by industry) of
- nonspecific methods and/or of assays with demonstrated insufficient quality.



If these limits are not objectively defined and fulfilled, there is a risk of letting error gain the upper hand, thus obscuring the clinical information supplied by the result and even causing negative effects on patients' outcome.

